Interaction of the *Bacillus stearothermophilus* ribosomal protein S15 with rRNA

by

Robert T. Batey

B.S. Biological Sciences
University of California, Irvine (1990)

B.S. Chemistry
University of California, Irvine (1990)

Submitted to the Department of Biology in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Biology

at the
Massachusetts Institute of Technology
February 1997

Signature of Author.....
Department of Biology
December 17, 1996

Certified by.....
James R. Williamson
Associate Professor of Chemistry
Thesis Supervisor

Accepted by.....
Richard Young
Chairman, Biology Graduate Committee

Interaction of the *Bacillus stearothermophilus* ribosomal protein S15 with rRNA

by

Robert T. Batey

Submitted to the Department of Biology in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Biology

ABSTRACT

The ribosomal RNA binding site of *Bacillus stearothermophilus* ribosomal protein S15 (BS15) was analyzed using synthetic RNA oligonucleotides derived from the 16S rRNA central domain. Native gel electrophoresis mobility shift assays demonstrate that BS15 can specifically interact with an RNA oligonucleotide containing nucleotides 585-756 (helices 20-23) of 16S rRNA with an apparent dissociation constant of 35 nM. A series of deletion mutants of the rRNA fragment that contains the BS15 specific binding site was tested for their capacity to bind protein using a competition binding assay, yielding a 61 nucleotide RNA that is capable of binding BS15 with wild type affinity. The major determinant of the BS15-rRNA interaction is a three-way junction between helices 20, 21, and 22, while helix 23 (nucleotides 673-733 of 16S rRNA) was dispensable for high affinity binding. Kinetic measurements of the dissociation and association rate of the bimolecular complex between BS15 and various minimal rRNA binding sites demonstrate that the kinetic and thermodynamic properties of this interaction were not altered as a result of the deletions. The sequence and structural determinants for the RNA in the BS15-rRNA interaction have been probed using site-directed mutagenesis, chemical modification interference, and iodine footprinting of phosphorothioate RNA. Mutations and RNA modifications that interfere with protein binding cluster in two distinct regions, one containing an internal loop and the other containing a three-way junction. Projection of this data onto a three-dimensional model illustrates that BS15 is likely to interact with the minor groove along an extended face of helix 22.

This minimized BS15-rRNA interaction has been examined using NMR spectroscopy. As a prelude to structural studies of this large protein-RNA complex by NMR, a general method for large scale preparation of uniformly isotopically labeled ribonucleotides and RNAs has been developed. For ^{15}N-labeling, *E. coli* is grown on ^{15}N-ammonium sulfate, whereas for ^{2}H- and ^{13}C-labeling, *Methylophilus methylotrophus* is grown on ^{13}C-methanol, which is more economical than ^{13}C-glucose. While the 30.5 kDa BS15-Fr15 complex yields high quality homonuclear spectra, indicative that this complex is well-behaved under NMR conditions, the introduction of ^{13}C into the RNA causes severe line broadening in heteronuclear 3D-NMR spectra, limiting the applicability of these types of experiments.

Thesis Supervisor: James R. Williamson

Title: Associate Professor of Chemistry

Acknowledgments

This thesis is not the accomplishment of a single person, but rather a synthesis of the ideas, experience, and passion of many people. While I reserve these few pages to express my appreciation for the support of many people, my ultimate gratitude is reflected in the pages that follow.

First, and foremost, I am indebted to Jamie. The S15 project was founded after he challenged me to go off and find an interesting RNA-protein interaction to study. His support for this project, even during the fabled dark year, has always been enthusiastic and wholehearted. During the course of this project, he has allowed me the intellectual and scientific freedom to explore this project, which I believe, more than anything else, has contributed to a rewarding graduate career. However, without his strong guidance and unyielding belief that this project *must* work, the S15 story would have been left on the cutting room floor years ago.

In the early years of my graduate career, the Williamson lab was graced by the presence of Jody Puglisi. There are not enough words to accurately describe the impact Jody had on molding the ideas and nature of this lab. Jody brought to this laboratory a working and practical knowledge of RNA biochemistry and NMR that profoundly impacted all of our studies (as evidenced by the fact that the most used document in this lab for the last five years has been his thesis). I hold the pearls of wisdom about RNA, science, and life in general he has given me over the years in the highest regard.

The people within the Williamson laboratory during my career have contributed much to this thesis. During my time here, I have had the pleasure to interact with a group of people that brought together an impressive array of skills. Pat Zarrinkar has provided many fruitful discussions about RNA techniques and biochemistry. Furthermore, without him, many of the day-to-day aspects of the lab would have fallen into complete chaos. While his vast organizational skills have provided us with many hours of amusement at his expense, there has not been a single person whose life in this lab has not been made a lot easier by his efforts. Similarly, Dan Treiber has made many positive contributions to this work with his strong working knowledge of molecular biology and protein-nucleic acid interactions. All of the NMR work in this thesis has been with the help and training of

John Battiste and Alex Brodsky. These guys have demonstrated a lot of patience helping me out with the technical aspects of NMR spectroscopy and interpreting the results.

A number of people in this lab have also directly collaborated with me in the S15 project. Maki Inada and Elizabeth Kujawinski, as UROP students, worked on developing techniques for the preparation of isotopically labeled nucleotides. This protocol has undergone considerable revision since its first version, in part due to the insightful observations and modifications of Tom Tolbert, whose spin on these methodologies helped to streamline the process. Normand Cloutier and Hongyuan Mao played strong facilitative roles in the improvements in methylotrophic growth that allowed us to grow these bacteria in deuterated media. Sharon Parker, in her brief stay as a postdoctoral associate in this lab, made significant contributions to our understanding of the structure (or lack thereof) of S15 in her work on the unbound form of *E. coli* and *B. stearothermophilus* S15. Her biochemical and NMR work elucidated properties of this protein that have greatly forwarded the studies presented in this thesis. This work has been supplemented more recently by Feng Tao. It has also been my pleasure to work with Jeff Orr on the S15 project in studies that have provided us with a global picture of the three-way junction, which has been invaluable in terms of stimulating our thinking about this complex. Also, special thanks to Jeff for his excellent editorial critiques and proofreading of large segments of this thesis.

Of course, I cannot forget Christopher Cilley. As a good friend and baymate, he has taught me a lot during this time. Although he has made many scientific contributions to this work, particularly in protein purification and general molecular biology techniques, and the administration of a minimally frustrating computer network, his most valuable contribution to this thesis has been non-scientific. He has done much to take the razor's edge off of my time at M.I.T. and has been consistently a valued friend that I could count upon.

I would also like to thank the contributions made by the M.I.T. Department of Biology, particularly Peter Kim and Carl Pabo as members of my thesis committee, and David Bartel and Lee Gehrke as additional members of my defense committee. Also, much of this work has been made possible by the M.I.T. Department of Chemistry and Division of Toxicology through fruitful interactions with professors and students in these departments, as well the equipment of Chemistry Department Spectroscopy Lab. Many hours of NMR time were also provided by the Francis Bitter National Magnet Laboratory, along with the assistance of Chris Turner in using their spectrometers.

In large part, I was guided to this graduate career by many people at the University of California, Irvine. My experience at UCI was truly great because of the many

wonderful professors, staff, and students I met in my years there. In particular, I would like to thank Prof. Donald Senear for allowing me to work as an undergraduate in his lab and stimulating my interest in protein-nucleic acid interactions. I am also indebted to Athena Spanoyannis, the lab technician in Don's lab, who gave me a great foundation in practical aspects of molecular biology. The other professor at UCI who had a strong influence in my education was Prof. David Brandt of the Department of Chemistry, who, without a doubt, is the best educator I have ever come across. Finally, I would like to thank Craig Longeauvan who provided excellent counsel during my time at UCI.

Of course, the greatest debt of gratitude is to my family. Any success I achieve is firmly predicated upon the environment in which I was raised. As an elementary school teacher, my mother has always strived to instill in us a firm respect and understanding of the power of solid education. In my younger years, she sacrificed much to spend the time to fill in the gaps and expand that education. From my father, I received the gift of the strong work ethic; above all, this I believe is the surest means to success.

Table of Contents

ACKNOWLEDGMENTS 3

TABLE OF CONTENTS 6

CHAPTER 1: INTRODUCTION 13

Ribonucleoprotein assemblages in biology 13
Genomic replication and transcription 13
mRNA processing and transport 16
Translation and translocation 17

The 16S rRNA 19
Organization of the 16S rRNA 19
Function of the 16S rRNA 25

Protein-RNA interactions in the 30S subunit 28
In vitro assembly of the prokaryotic 30S subunit 29
RNA-protein interactions within the ribosome 31
Towards the structure of the 30S subunit 34

The ribosomal protein S15 37
Interaction with 16S rRNA 39
S15 as a repressor of its own translation 41

CHAPTER 2: METHODS FOR ANALYSIS OF RNA STRUCTURE AND RNA-PROTEIN INTERACTIONS. 44

RNA sequencing and secondary structure mapping 44
Nuclease sequencing and mapping 44
Chemical sequencing 48

Gel Electrophoresis Mobility Shift Assay 57
Determination of the apparent equilibrium dissociation constant 60
Determining the stoichiometry of the protein-nucleic acid interaction 62
Determining the relative binding constants of RNAs 62
Kinetics 64
RNA folding 66

Chemical modification interference analysis 66

Chemical Footprinting 77

The dimerization question 80

Nuclear Magnetic Resonance (NMR) Spectroscopy 83
NMR applied to RNA 84
Chemical exchange and relaxation processes in NMR 91
Sample preparation 94
NMR spectroscopy 96

CHAPTER 3: DETERMINATION OF THE MINIMAL RRNA BINDING SITE FOR S15 98

Interaction of BS15 protein with Fr1 RNA 98

Systematic deletion of Fr1 RNA to produce a minimal binding site for BS15 104

Kinetics of the BS15-RNA interaction 113

Discussion 117

CHAPTER 4: ANALYSIS OF THE S15 BINDING SITE IN RRNA 123

Site specific mutations in Fr1 RNA 123

Identification of important nucleotides in Fr1 for BS15 binding 127

Interference analysis of BS15-Fr7 RNA interaction 131

Detailed analysis of the S15-rRNA determinants 136

Footprinting of the BS15-Fr15 interaction 147

Discussion 152
Interpretation of the determinants of the BS15-rRNA interaction 152
The conformation of the rRNA three-way junction 158
Relationship between the rRNA and mRNA binding site of S15 160

CHAPTER 5: FOLDING OF THE RNA THREE-WAY JUNCTION 163

Nucleic acid junctions 163

Folding of Fr1 RNA and mutants in the presence of polyvalent ions 166

Modification interference of the RNA conformational change 170

CHAPTER 6: NMR STUDIES OF THE BS15-RRNA INTERACTION 178

NMR spectroscopy of the BS15-rRNA complex 178
Quality of RNA and protein utilized for NMR studies 178
Formation of the BS15-rRNA complex in the NMR tube 180

Optimization of conditions for NMR spectroscopy 182
Heteronuclear spectroscopy of the BS15-Fr15 complex 194
Future directions for structural studies of the BS15-rRNA complex 196

NMR studies of a minimal three-way junction element 198

CHAPTER 7: FUTURE DIRECTIONS 204

Further RNA studies 204

Towards the solution structure of S15 210

Assembly of an RNP particle comprising the 30S central domain 213

APPENDIX 1: SYNTHESIS AND PURIFICATION OF BIOLOGICAL MATERIALS 215

Expression and purification of proteins 215

***Bacillus stearothermophilus* protein S15 216**
Construction of the BS15 overexpression plasmid 216
Purification of the BS15 Protein 220

***Escherichia coli* ribosomal protein S15 223**
PCR amplification of the protein from *E. coli* genomic DNA 223
Expression controls 224
Purification 225

T7 RNA Polymerase 225

Synthesis and Purification of RNAs 227
Purification of *Bacillus stearothermophilus* 16S rRNA 227
Construction of plasmids containing RNA genes 228
Preparation of Fr1 RNA mutants 231
Synthesis and purification of RNA substrates 231
Synthesis of RNA from DNA oligonucleotides templates 234

APPENDIX 2: PREPARATION OF RNAS LABELED WITH STABLE NUCLEI FOR NMR ANALYSIS 236

Introduction 236

Methodology 239
Growth of *E. coli* 239
Growth of *Methylophilus methylotrophus* 240
Cellular Lysis 247
Nucleic acid hydrolysis 248
Nucleotide separation 249
Separation of individual ribonucleotides 252
Enzymatic phosphorylation 252
Nucleotide purification 256

In Vitro Transcription of TAR RNA 256

Practical Aspects **258**
Growth of *E. coli* 258
Growth of *Methylophilus methylotrophus* 260
Cellular Lysis 263
Nucleic acid hydrolysis 265
Nucleotide separation 265
Separation of individual ribonucleotides 267
Enzymatic phosphorylation 268
Nucleotide purification 272
In Vitro Transcription 272

NMR of u-^{13}C-labeled TAR **273**

BIBLIOGRAPHY **277**

BIOGRAPHY **303**

Table of Figures

Figure 1.1: Central role of RNA in the cell 14
Figure 1.2: Secondary structure of the *E. coli* 16 S rRNA 20
Figure 1.3: Phylogenetic conservation of nucleotides in the 16S rRNA 21
Figure 1.4: Secondary structural motifs of the 16S rRNA 22
Figure 1.5: Functional domains of the 16S rRNA 26
Figure 1.6: Assembly pathway of the 30S ribosomal subunit 30
Figure 1.7: Model of the 30S ribosomal subunit 35
Figure 1.8: Sequences of phylogenetic variants of ribosomal protein S15 38
Figure 1.9: Summary of biochemical studies of the *E. coli* S15-rRNA interaction 40
Figure 1.10: Interaction of S15 with an mRNA translational operator site 42

Figure 2.1: DMS modification of the guanine base 50
Figure 2.2: DEPC modification of RNA 52
Figure 2.3: Hydrazine modification of RNA 53
Figure 2.4: Aniline mediated strand scission of RNA 55
Figure 2.5: Aniline modification of RNA 56
Figure 2.6: Modification interference analysis scheme 67
Figure 2.7: Ethylnitrosourea modification of RNA 71
Figure 2.8: DMS modification of RNA bases 74
Figure 2.9: CMCT modification of RNA bases 75
Figure 2.10: Iodine mediate strand scission of phosphorothioate containing RNA 79
Figure 2.11: Size exclusion chromatography of RNA 82
Figure 2.12: Exchangeable and nonexchangeable protons in RNA 85
Figure 2.13: Nonexchangeable proton NMR spectrum of RNA 86
Figure 2.14: Exchangeable proton NMR spectrum of RNA 88
Figure 2.15: Conformations of the ribose sugar in RNA 90
Figure 2.16: Effect of exchange in NMR spectroscopy 92

Figure 3.1: Direct titration of Fr1 RNA with BS15 100
Figure 3.2: Stoichiometry of the BS15-Fr1 RNA interaction 103
Figure 3.3: The family of deletion mutant RNAs 105
Figure 3.4: Gel mobility shift of representative deletion mutants 108
Figure 3.5: Competition assays using deletion mutant RNAs 110
Figure 3.6: Dissociation kinetics of the BS15-Fr1 RNA interaction 115
Figure 3.7: Association kinetics of the BS15-Fr1 RNA interaction 118

Figure 4.1: Mutations introduced into Fr1 RNA 124
Figure 4.2: Modification interference assays of the BS15-Fr1 RNA interaction 129
Figure 4.3: Modification interference assays of the BS15-Fr1(U653Δ) interaction 132
Figure 4.4: Modification interference assays of the BS15-Fr7 interaction 134
Figure 4.5: Modification interference assays of the BS15-Fr15 interaction 139
Figure 4.6: ENU modification interference assays of the BS15-Fr15 interaction 143
Figure 4.7: Modification interference assays of the BS15-Fr15/*Eae*I interaction 145
Figure 4.8: Base pairing interactions in the three-way junction element 148
Figure 4.9: Iodine footprinting of the BS15-Fr15 interaction 150
Figure 4.10: Comparison of 16S rRNA phylogeny and modification interference data 153
Figure 4.11: Model of the S15-rRNA interaction 156
Figure 4.12: Conformational change in the 16S rRNA mediated by S15 159
Figure 4.13: Comparison of the rRNA and mRNA binding sites of S15 161

Figure 5.1: Migration of Fr1 RNA and mutants in a native gel with EDTA and Mg^{2+}....167
Figure 5.2: Migration of Fr1 RNA and mutants in a native gel with Co^{3+}.....169
Figure 5.3: Migration of Fr1 RNA and mutants in a native gel with spermidine.....171
Figure 5.4: Modification interference of Mg^{2+}-dependent folding of Fr4 RNA.....173
Figure 5.5: Conformational change of 16S rRNA mediated by divalent ions.....177

Figure 6.1: Imino titration of Fr15 RNA with BS15.....181
Figure 6.2: Temperature dependence of the NMR spectra of the BS15-Fr15 complex...183
Figure 6.3: Analysis of Fr15 RNA stability in the BS15-Fr15 complex.....187
Figure 6.4: DQF-COSY spectra of Fr15 RNA in the absence and presence of BS15.....189
Figure 6.5: 750 MHz spectrum of the BS15-Fr15 RNA complex.....191
Figure 6.6: CT-HSQC spectrum of ^{13}C-guanosine labeled BS15-Fr15 RNA complex..195
Figure 6.7: Fr18 RNA.....199
Figure 6.6: Imino titration of Fr18 RNA with magnesium.....201
Figure 6.7: DQF-COSY spectra of Fr18 RNA with magnesium.....202

Figure 7.1: S15-binding tripartite RNA complex.....206
Figure 7.2: Formation of the RNA tripartite complex.....207
Figure 7.3: ^{15}N-HMQC spectra of free and bound BS15 protein.....211

Figure A1.1: Construction and map of pEXBS15 plasmid.....218
Figure A1.2: Sequence of the synthetic BS15 gene.....219
Figure A1.3: Purification gel of BS15 protein.....222
Figure A1.4: Optimization of expression of ES15 protein.....226
Figure A1.5: Plasmid constructs for templates for RNA transcription.....229
Figure A1.6: Synthesis of RNA from single stranded templates.....235

Figure A2.1: Flowchart of the protocol for labeling RNA with stable isotopes.....238
Figure A2.2: Growth of *Methylophilus methylotrophus*.....244
Figure A2.3: Growth of *Methylophilus methylotrophus* with pH control.....246
Figure A2.4: Nucleotide composition of *Methylophilus methylotrophus*.....250
Figure A2.5: Boronate separation of rNMPs from dNMPs.....251
Figure A2.6: Nucleotide composition of boronate chromatography fractions.....253
Figure A2.7: Separation of the individual ribonucleotide monophosphates.....254
Figure A2.8: Phosphorylation of rNMPs to rNTPs.....257
Figure A2.9: Transcription of TAR RNA with labeled rNTPs.....259
Figure A2.10: Optimization of methanol utilization.....262
Figure A2.11: Reaction of immobilized boronate ligand with ribose.....266
Figure A2.12: ^{1}H spectrum of ^{13}C-labeled CMP.....269
Figure A2.13: ^{1}H and ^{13}C spectrum of deuterated UMP.....270
Figure A2.14: Chemistry of the phosphorylation reaction.....271
Figure A2.15: HCCH-COSY spectrum of u-^{13}C-TAR RNA.....274

Table of Tables

Table 2.1: Nucleases and chemicals used for RNA sequencing and probing.....45

Table 3.1: Apparent dissociation constants for the BS15-Fr1 interaction.....101
Table 3.2: Binding data of deletion mutants.....112
Table 3.3: Temperature dependence of k_{off}.....116

Table 4.1: Binding data of Fr1 point mutants.....125
Table 4.2: Modification interference and footprinting data for Fr15 RNA.....137

Table 6.1: Temperature dependence of T_2 relaxation time of the BS15-Fr15 complex....184

Table A2.1: Media used for growing bacteria on isotopically labeled substrates.....241
Table A2.2: Yields of rNMPs from *E. coli* and *M. methylotrophus*.....264

Chapter 1: Introduction

Ribonucleoprotein assemblages in biology

The flow of information in cellular life is shepherded by a series of ribonucleoprotein (RNP) assemblages whose function is to maintain the integrity of the genome and the efficient transfer of its information into functional RNAs and proteins. The current view of the RNA components of these assemblages as vital and active participants in their functions was galvanized by the discovery of a class of RNAs that was capable of chemical catalysis independently of proteins (Kruger *et al.*, 1982; Guerrier-Takada *et al.*, 1983). From this discovery arose the hypothesis of the "RNA world", an era of early life in which RNA exclusively performed the functions of genome and enzyme (Gilbert, 1986). However, this RNA world, if it existed, gave rise to the current DNA/RNA/protein world in which RNA plays a critical role translating the genetic information of DNA into the catalytic functions of proteins. For the most part, the role of RNA in cellular processes is enhanced by the participation of proteins that specifically recognize, bind, and function as an inseparable part of the ribonucleoprotein assemblage. Along the entire pathway from DNA gene to functional protein, ribonucleoprotein particles play an active role in the management of information transfer (Figure 1.1).

Genomic replication and transcription

Although cellular life has abandoned the use of RNA as the repository of inheritable genetic information, it still plays an integral role in the replication of DNA and RNA genomes. RNA does serve as the repository of the genome in retroviruses such as the human immunodeficiency virus (HIV). During later stages of infection, genomic RNA

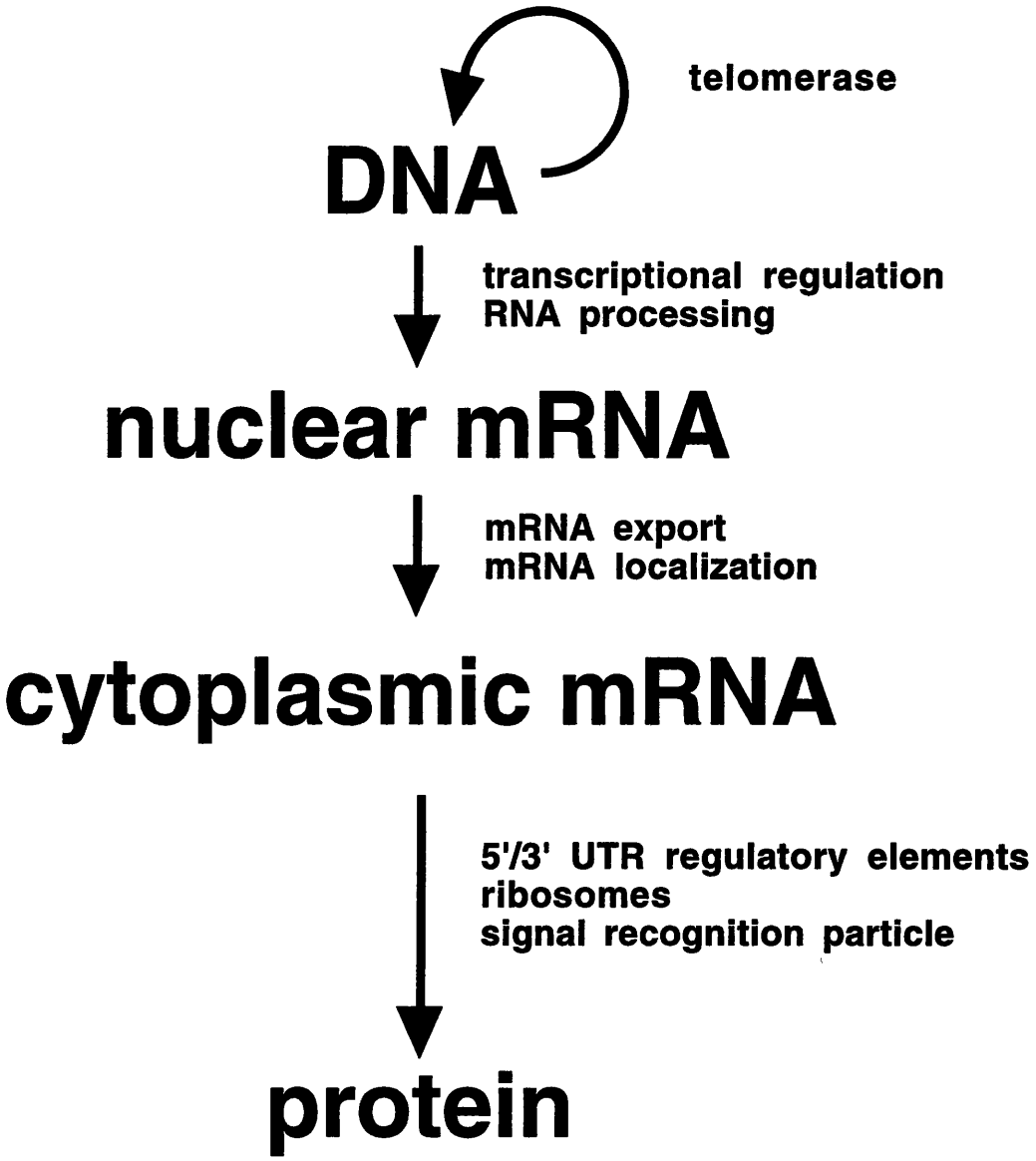


Figure 1.1: The central role of ribonucleoprotein assemblages in information transfer in the cell.

must be efficiently transcribed from the viral genome that is incorporated into the host genome. This is in part accomplished by the regulation of the transcriptional assembly via its interaction with the *trans*-activating region (TAR) found in the 5' long terminal repeat (LTR) of the transcribed RNA (Cullen & Malim, 1991; Karn, 1991). The regulatory function of the TAR element of the RNA is mediated by its interaction with the virally encoded protein Tat to greatly enhance the processivity of transcription of the viral RNA (Frankel, 1992; Greenbaum, 1996). A similar mechanism for enhancing transcriptional processivity is used by a number of bacteriophages. In order for bacteriophage λ to express the late phase genes, it must enable the transcriptional complex to transcribe through a series of strong terminator elements in the mRNA (Greenblatt *et al.*, 1993). This is accomplished by the interaction of the N protein with an RNA element, box B, which then interacts with RNA polymerase to allow it to override the transcription terminator sequences.

Retrotransposable elements, believed to be relatives of the retroviruses, also integrate a DNA element that was reverse transcribed from an RNA into a host genome (reviewed in Curcio & Belfort, 1996). The mobility of a group II intron, one type of self-splicing mRNA intron, has recently been shown to insert itself into DNA via a unique mechanism (Zimmerly *et al.*, 1995). The DNA endonuclease activity associated with this retroelement is composed of the group II intron and an associated protein. The RNA component of this endonuclease cleaves one strand of the DNA, while the protein component has been shown to cleave the other strand. This is a clear example of the catalytic activity of a ribonucleoprotein enzyme divided between the RNA and protein components, each performing very similar functions.

The reverse transcriptase activity that enables retroviruses and retroelements to convert their RNA genome into DNA that can be inserted into a host genome is not unique to these entities. Telomerase, an enzyme that is responsible for maintaining the proper length of the terminal sequences of linear chromosomes, the telomeres, is also a reverse

transcriptase (Blackburn, 1993). Embedded within the RNA component of this ribonucleoprotein enzyme, there is a sequence that serves as the template for telomeric DNA synthesis. Without the action of telomerase, the linear ends of the chromosome could not be replicated, resulting in chromosomal shortening with each cycle of replication, ultimately resulting in cell death (Zakian, 1995).

mRNA processing and transport

Within the nucleus of eukaryotic cells are a myriad of small nuclear ribonucleoprotein particles (snRNPs) responsible for the processing of mRNA, rRNA, and tRNA. Newly transcribed mRNAs reflect the exon/intron structure of the DNA genome, and it is the responsibility of the spliceosome to excise the intron sequences to yield an mRNA with a proper coding sequence. The spliceosome is a complex assemblage of the snRNPs U1, U2, U4/U6, and U5, each containing a small nuclear RNA (snRNA) and a number of accessory proteins (Moore *et al.*, 1993). These snRNPs interact with each other and the mRNA substrate to varying degrees and at different times during the splicing process to create a functional spliceosome. Similarities between the chemistry of splicing and the structure of the RNA components of the spliceosome and the group II intron has led to the suggestion that the catalytic activity of the spliceosome resides within the RNA components (Guthrie, 1991; Madhani & Guthrie, 1994). Nonetheless, the protein components of this ribonucleoprotein assemblage play an integral role in the splicing process along with other roles such as the regulation of splicing and splice site selection.

RNase P is an essential ribonucleoprotein enzyme that produces mature tRNA from primary tRNA transcripts by removal of the 5' leader sequence. In prokaryotes, the RNA component of this RNP has been shown to be catalytically active in the absence of the protein, although it requires the presence of high salt concentrations (Guerrier-Takada *et al.*, 1983; Guerrier-Takada & Altman, 1984). However, it is clear that the RNA is a far

more efficient and versatile enzyme in the presence of the protein (Altman *et al.*, 1993). In eukaryotes, the protein component of this enzyme is essential for activity, reflecting the critical role of the protein for modulating and enhancing the activity of RNA. Similarly, ribosomal RNA processing in the nucleolus of eukaryotes accomplished with the aid of the U3, U8, U13, U18, X, and Y snRNPs (Baserga & Steitz, 1993).

Incompletely spliced mRNAs are restricted to the nucleus through their interaction with the splicing machinery and other nuclear factors that likely recognize the splice sites (Legrain & Rosbash, 1989). However, lentiviruses have developed a mechanism that is capable of bypassing the normal routes of mRNA export to the cytoplasm in order to export late-phase unspliced mRNAs. In HIV, this is accomplished by the essential protein Rev, which binds to the Rev response element (RRE) within the mRNA and facilitates the export of incompletely spliced RNA out of the nucleus (Meyer & Malim, 1994). This is accomplished by the interaction of Rev protein with the cellular protein export pathway in the nucleus, allowing for incompletely spliced viral RNAs to escape the nucleus (Fritz & Green, 1996). Thus, the formation of mRNPs with a number of nuclear factors ultimately dictates the mechanism by which the mRNA is exported to the cytoplasm.

Translation and translocation

Once the mRNA has been transported from the nucleus into the cytoplasm, it interacts with another set of RNA binding proteins and RNPs that are responsible for proper localization of the mRNA in the cytoplasm, translation of its message, and targeting that protein to its proper destination (reviewed in Dreyfuss *et al.*, 1996). The first two tasks are accomplished with the help of a series of proteins that interact with the 5' and 3' untranslated regions (UTRs) of the mRNA. Upon export from the nucleus, the 5' N7-methyl guanylate cap interacts with a complex of proteins, eIF-4F, which are responsible for the regulation of translation via the cap-dependent mechanism. Assemblages associated

with the polyadenylated 3' end of mRNAs also play similar roles in the regulation of translation. These mRNPs have also been implicated in the regulation of mRNA stability and degradation (Cohen, 1995; Caponigro & Parker, 1996).

Translation of cytoplasmic mRNA into a protein is mediated by the ribosome, a complex macromolecular assemblage consisting of two RNPs, the small and large subunits. The small subunit (the 30S subunit in prokaryotes or the 40S subunit in eukaryotes) provides the half of the ribosome responsible for binding mRNA, formation of a preinitiation complex, the association of the mRNA codon with the tRNA anticodon at the decoding site, and association with translation factors such as the initiation factors and EF-Tu (comprehensively reviewed in Noller & Nomura, 1987). The large subunit (the 50S subunit in prokaryotes or the 60S subunit in eukaryotes) harbors the peptidyl transferase center, the site of synthesis of the elongating peptide, the EF-G binding domain, and the GTPase center. These functions are divided between "RNA" centers, such as the decoding site and peptidyl transferase center, whose activity is mediated by RNA-rich domains, and "ribonucleoprotein" centers such as the EF-Tu and EF-G binding sites, whose activities are mediated by both RNA and proteins (Noller, 1993).

Newly synthesized proteins also need to be targeted for intracellular or extracellular localization. This occurs during the early phase of synthesis of the protein with the aid of a small RNP called the signal recognition particle (SRP) that specifically recognizes a signal sequence at the amino terminus of a protein being actively translated and targets it for secretion into the endoplasmic reticulum (comprehensively reviewed in Lutcke, 1995). The eukaryotic SRP is composed of two functional RNP domains: the elongation arrest domain which temporarily blocks the elongation of the peptide until the ribosome can dock with the endoplasmic reticulum, and the signal recognition domain, which recognizes the amino terminal signal sequences of the protein. A homologous particle that is responsible for extracellular secretion of proteins is present in eubacteria and consists of a small portion of the signal recognition domain of the eukaryotic SRP (Poritz *et al.*, 1990).

Of all of these ribonucleoprotein assemblages, the 30S ribosomal subunit of the prokaryotic ribosome is the most extensively characterized. This RNP consists of a 16S rRNA, typically about 1540 nucleotides in length, and twenty-one individual proteins called r-proteins. This subunit, along with the less well characterized 50S subunit forms a ribonucleoprotein enzyme that is essentially a polypeptide polymerase.

The 16S rRNA

The small ribosomal subunit of eubacteria, archae, and plant chloroplasts all contain a single highly conserved RNA of approximately 1540 nucleotides in length. Using a combination of chemical and enzymatic probing techniques, and comparative analysis, a detailed secondary structure of the 16S rRNA has been developed (Noller & Woese, 1981; Gutell *et al.*, 1994). The current model of the secondary structure of *Escherichia coli* is shown in Figure 1.2. The 16S rRNA is comprised of three domains: the 5' domain (residues 1-556), a central domain (residues 564-912), and a 3' domain consisting of a major domain (residues 926-1391) and a minor domain (residues 1392-1542) (Noller & Woese, 1981). This domain structure is strictly conserved across all the 16S and 16S-like rRNA, and is also observed in the analogous RNA from the eukaryotic small ribosomal subunit 18S rRNA. Comparative analysis of over 2000 16S rRNA sequences has revealed that a number of sequences and structural elements within this RNA are conserved (Figure 1.3), which are undoubtedly critical for its organization and function.

Organization of the 16S rRNA

The structure of the 16S rRNA is composed of a variety of secondary structural motifs, some of which are illustrated in Figure 1.4. Over fifty double-stranded helical

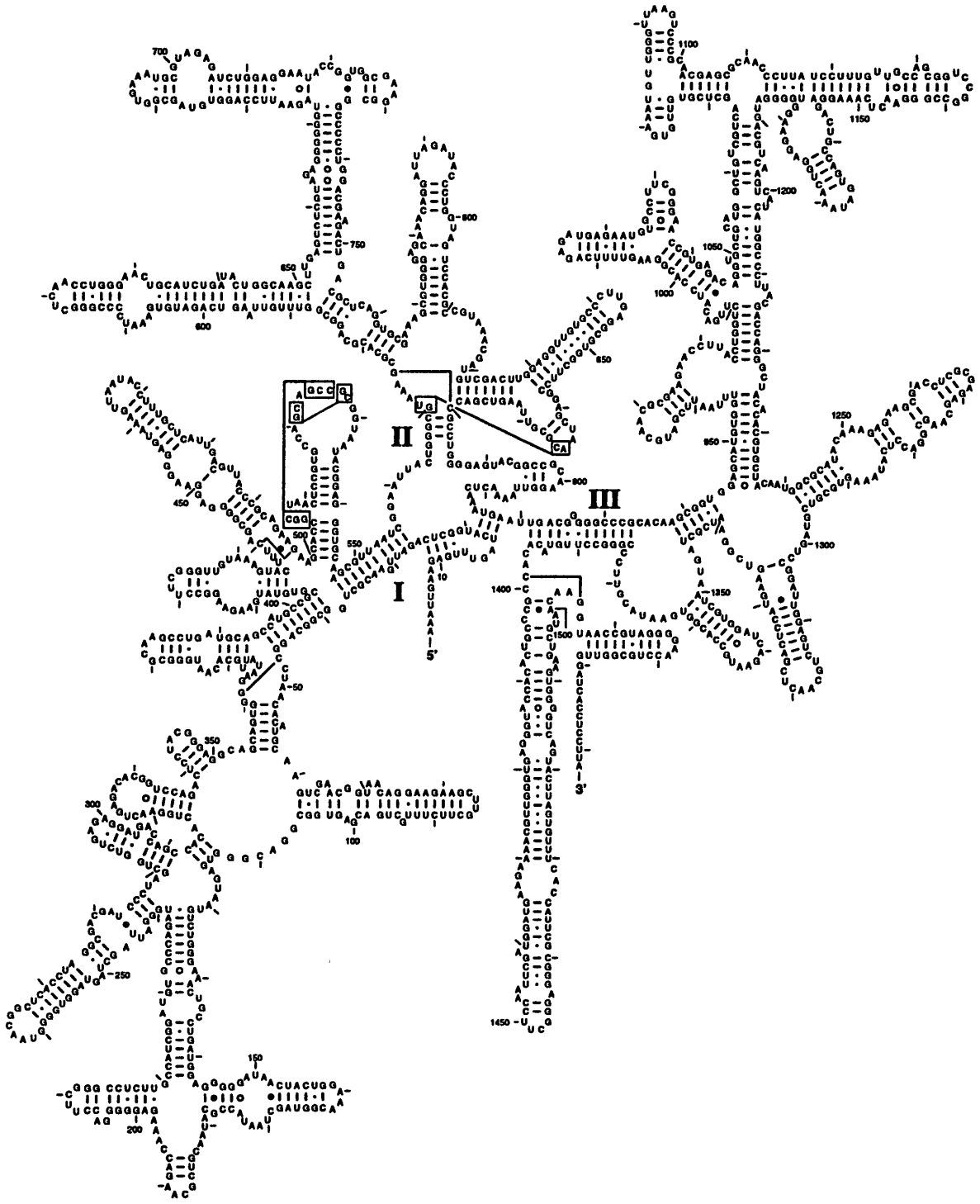


Figure 1.2: Sequence and secondary structure of the *E. coli* 16S rRNA. Adapted from Gutell, 1994.

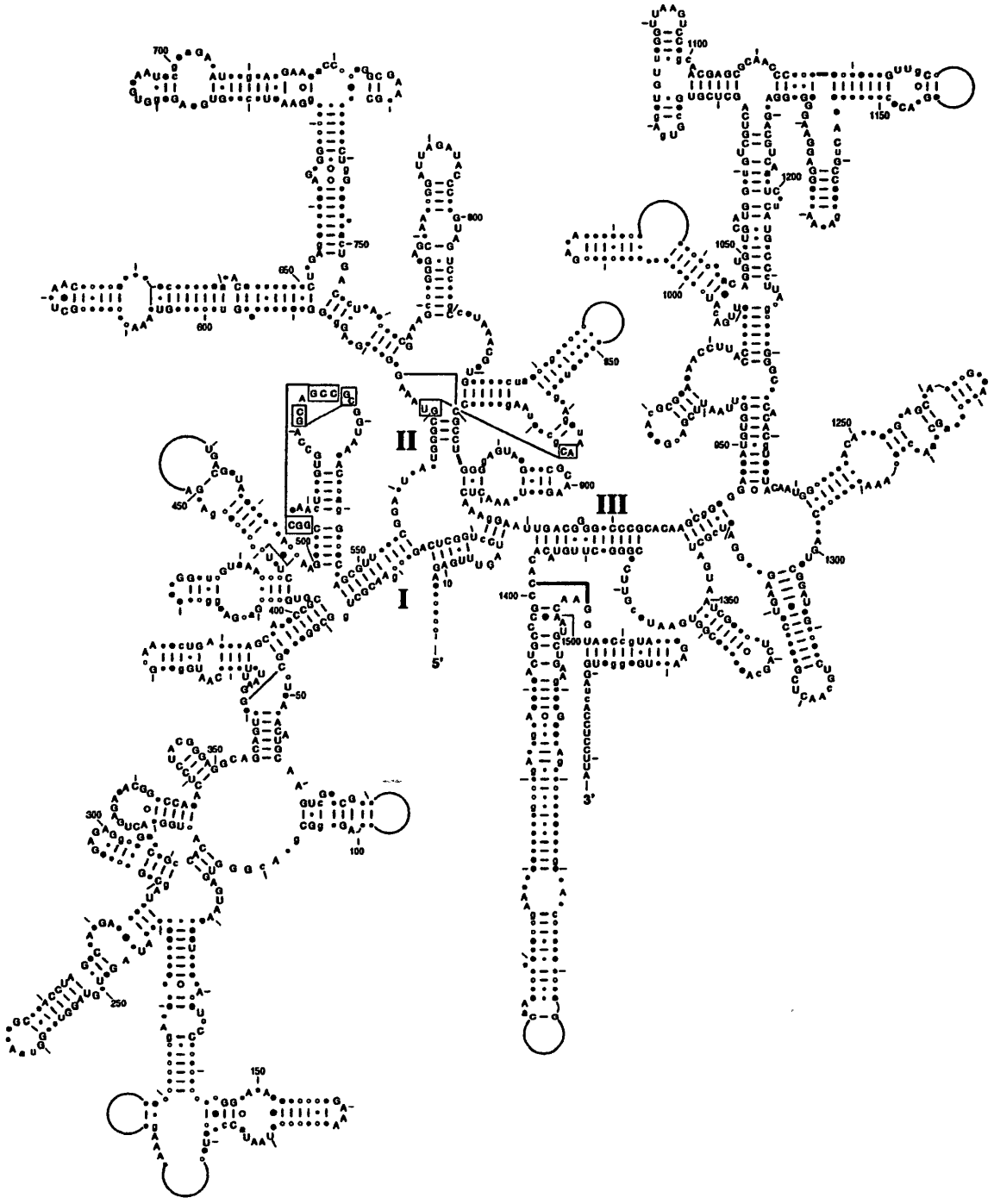


Figure 1.3: Phylogenetic conservation of nucleotides in eubacterial and chloroplast ribosomal 16S rRNA superimposed onto the *Escherichia coli* 16S rRNA structure. Captial letters denote positions that are >95% conserved in nucleotide identity, lower case letters represent positions that are 90-95% conserved, filled circles represent positions that are 80-90% conserved, and open circles represent nucleotides that exist relative to *E. coli* 95% of the time. Regions of the 16S rRNA that are highly variable in size and content are shown as arcs. Adapted from Gutell, 1994.

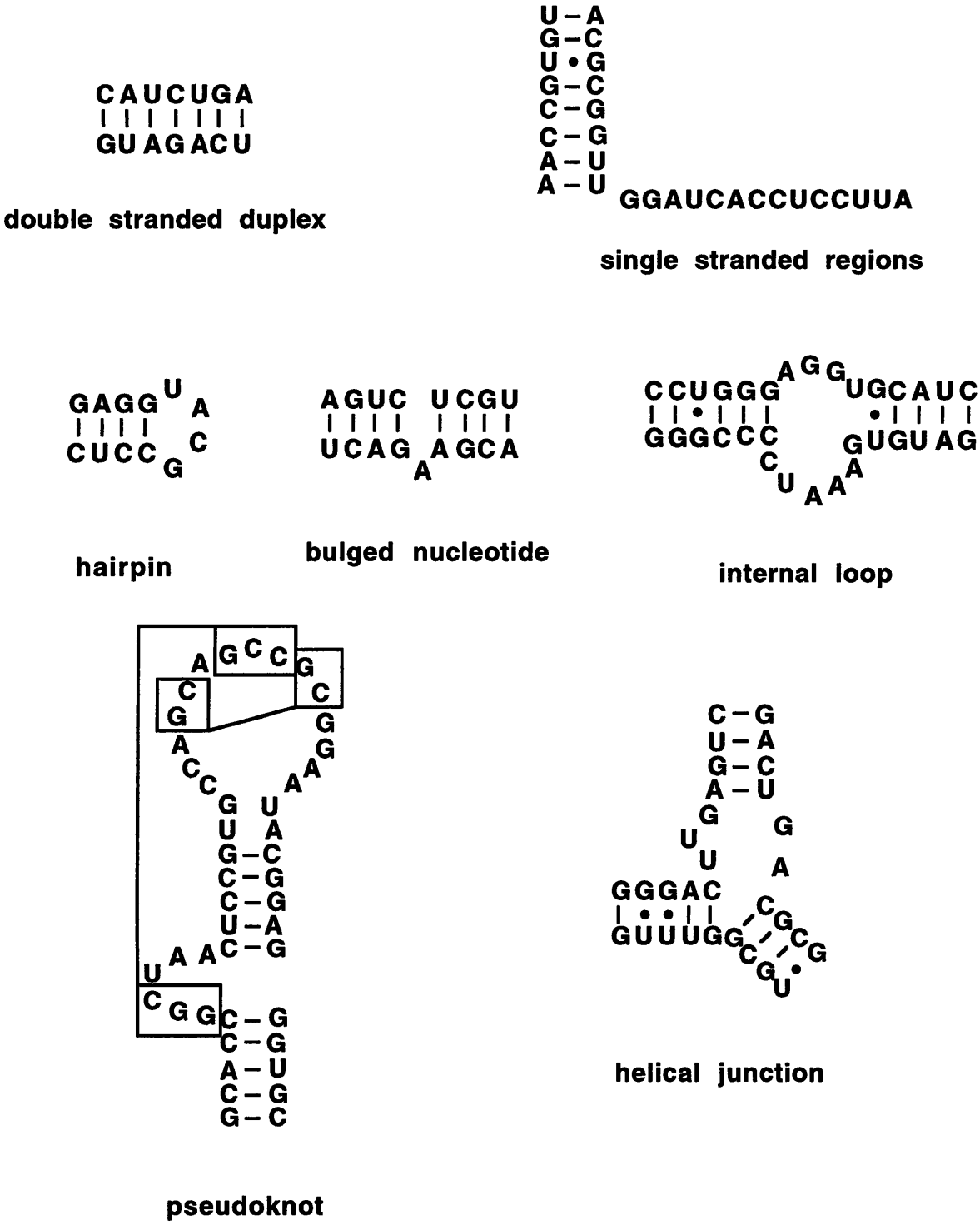


Figure 1.4: Structural elements commonly found in 16S rRNA.

elements define the three major domains. These helices are typically very short; not a single helix in 16S rRNA exceeds eight uninterrupted Watson-Crick G-C or A-U base pairs. Furthermore, while the presence of many of these helices are conserved across phylogeny, the identity of the nucleotides are often not. In this RNA, there is a distinct tendency for highly conserved nucleotides to fall outside the Watson-Crick paired helical elements (Gutell *et al.*, 1985). The incorporation of non-canonical base pairs (base pairs that do not have standard Watson-Crick A-U or G-C pairing geometries) are frequently found within the helical segments. In many of these cases, G-U pairs are often interchangeable with Watson-Crick pairs in phylogeny, however, there are many positions within the 16S rRNA where G-U pairs are invariant across phylogeny, possibly playing roles that can not be fulfilled by Watson-Crick pairs (Gautheret *et al.*, 1995). Another extensively utilized pairing type in 16S rRNA is the G•A pair, including several sites of multiple, contiguous pairs of this type. These G•A pairs demonstrate different patterns of replacement in phylogeny, suggesting that their geometries are not the same (Gutell, 1994). Other non-canonical pairings are not common to this RNA, but there are sites where G•G to A•A covariation exists, as well as conservation of a U•U pair.

Paired helical motifs are also broken by the frequent use of internal nucleotide bulges and loops. Within 16S rRNA, there are a number of sites that contain a single phylogenetically conserved bulged nucleotide. The role of these bulges in RNA is hard to determine since various studies have suggested that they can either loop out of the helix or intercalate into the helix to create a bend at that site (reviewed in Chastain & Tinoco, 1991). Studies on an identical DNA fragment with a single bulged adenosine yielded conflicting structures in which the adenosine was bulged out creating a straight helix using x-ray crystallography (Joshua-Tor *et al.*, 1988; Miller *et al.*, 1988) or a bent helix with the adenosine stacked in using NMR spectroscopy (Patel *et al.*, 1982; Roy *et al.*, 1987). Internal loops are similarly difficult to interpret. Within the 16S rRNA, there are many internal loops which contain highly conserved nucleotides whose roles are unclear. One

internal loop whose structure is well determined by NMR is the loop E motif of the 5S rRNA (Wimberly *et al.*, 1993; Szewczak & Moore, 1995). While this motif is still referred to as an internal loop, this RNA element, as well as others of the same class, contain several non-canonical base pairs and a single unpaired nucleotide bulged out of the helix. Since there appears to be some variability in the sequences that can form this motif, it may be hard to predict the pairing scheme of a given internal loop. Thus, many of the other internal loops in 16S rRNA that are currently drawn as unpaired may, in fact, form double stranded helical structures containing non-canonically paired bases. More recently, it was determined that an internal loop found in group I and group II introns is important for the tertiary organization of the RNA (Jaeger *et al.*, 1994). In this case, a GAAA tetraloop motif (see below) docks into an internal loop motif to create a tertiary interaction important for achieving a stably folded structure (Cate *et al.*, 1996). This principle will likely apply to the organization of the 16S rRNA as well.

One of the first structural motifs to be recognized in 16S rRNA was the tetraloop motif (Woese *et al.*, 1990). This motif is widespread throughout the 16S rRNA and is divided into three families: GNRA, UNCG, and CUUG. The structures of these three motifs have been solved by NMR, and in each case, the nucleotides at the one and four positions of the tetraloop form a non-canonical pair and the third nucleotide is stacked on top of the fourth nucleotide (Cheong *et al.*, 1990; Heus & Pardi, 1991; Jucker & Pardi, 1995). Initially, it was believed that these motifs were predominantly found in RNA because of their very high stability. More recently, in crystal structures of the hammerhead ribozyme and the P4/P6 domain of the group I intron, the GNRA tetraloop motif has been found to make tertiary contacts to other parts of the RNA (Pley *et al.*, 1994a; Cate *et al.*, 1996). Thus, tetraloop motifs appear to also play a role in the organization and folding of larger RNAs.

Along with the tetraloops, several other structural motifs are represented in the 16S rRNA. The pseudoknot is a frequently observed motif in ribosomal RNA. The most

phylogenetically conserved of all elements in the 16S rRNA, found in the 500-540 region, has been shown to form a pseudoknot (Gutell & Woese, 1990; Powers & Noller, 1991). As will be discussed later, this pseudoknot plays a critical role in the function of the 16S rRNA. Another pervasive motif in this RNA is the helical junction; three, four, and five-way junctions are found throughout the RNA. Again, these motifs tend to harbor conserved nucleotides, particularly in the unpaired elements found at the helical junction itself. One of the fundamental questions concerning the global architecture of the 16S rRNA is how helices are arranged with respect to one another. For adjacent helices, it is believed that there is a strong driving force for the helices to coaxially stack. While the structure of pseudoknot motifs has been elucidated in various NMR studies (Puglisi *et al.*, 1990; Shen & Tinoco, 1995), the junction motif and the rules for how helices coaxially stack and orient themselves in some of the more complicated junction motifs found in the 16S rRNA are not well understood.

Function of the 16S rRNA

The relatively recent discovery of a number of RNAs that are capable of chemical catalysis has been the impetus for thinking that RNA plays a fundamental role in the function of ribonucleoprotein assemblages, including the ribosome. However, long before the discovery of RNA enzymes (ribozymes), it was suggested that the ribosomal RNA is more than a bit player in ribosomal activity. This was highlighted by early experiments in which is was found that the antibiotic colicin E3 specifically inactivated the 30S subunit by specifically modifying the 16S rRNA, destroying the ability of the ribosome to initiate protein synthesis (Nomura, 1963; Konisky & Nomura, 1967). Since then, specific functions have been ascribed to many parts of the 16S rRNA (Figure 1.5).

The first function for 16S rRNA to be elucidated came as a result of the colicin E3 studies. Using colicin E3 cleavage as an assay, the base pairing interaction between a

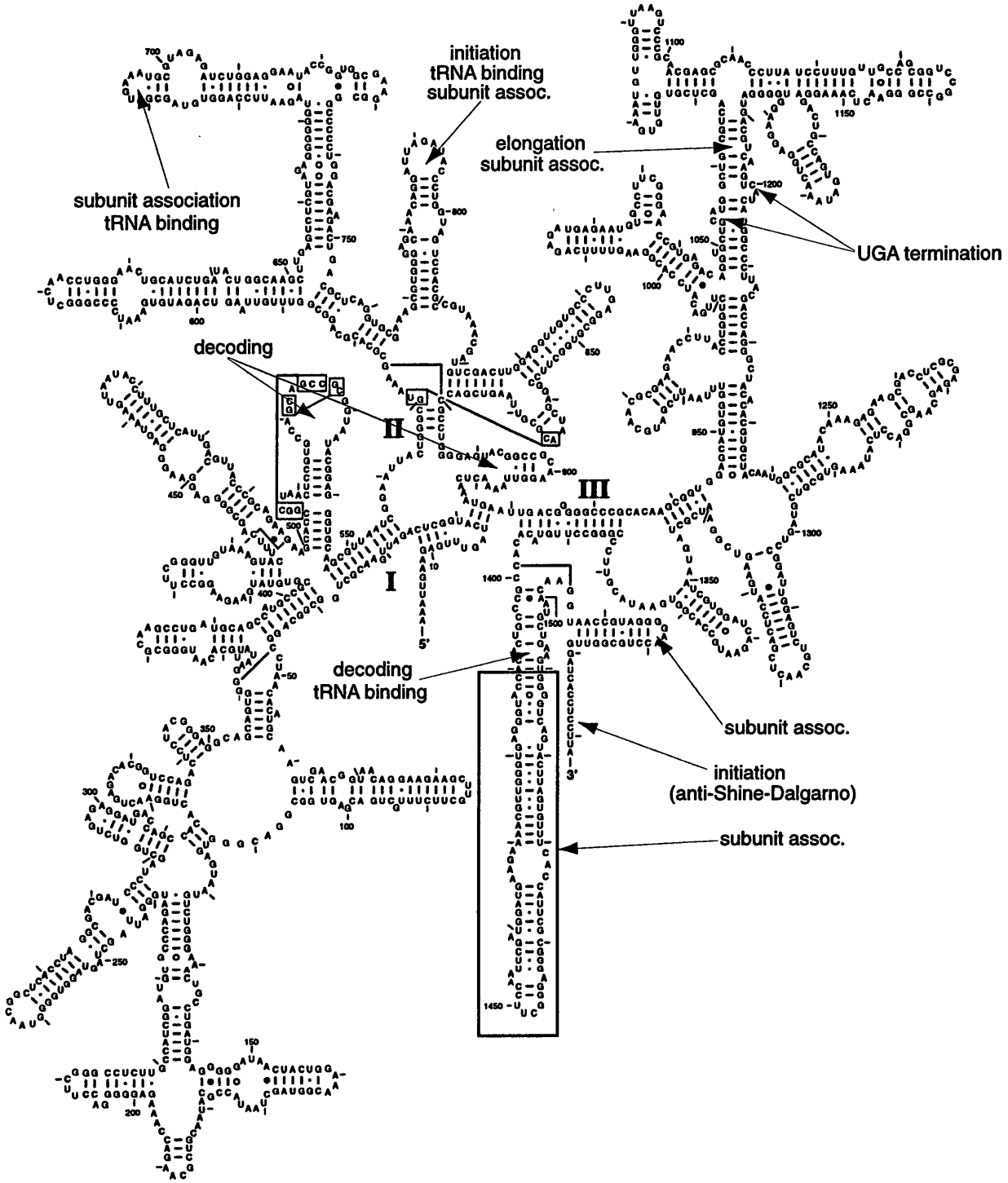


Figure 1.5: Functional regions in the 16S rRNA superimposed upon the *E. coli* 16S rRNA structure. Adapted from Raue *et al.*, 1990.

sequence in the mRNA immediately upstream of the initiator AUG codon, the Shine-Dalgarno sequence (Shine & Dalgarno, 1974), and a single stranded sequence at the extreme 3' end of the 16S rRNA was demonstrated (Steitz & Jakes, 1975). Mutations in the anti-Shine-Dalgarno in the 16S rRNA or the Shine-Dalgarno sequence in the mRNA that reduce base pairing reduces translation, while compensatory mutations that increase base pairing increase translation (Jacob *et al.*, 1987). mRNA-rRNA interactions are also found to affect elongation and termination phases of transcription. During elongation, the ribosome is known to shift reading frames at specific points in the mRNA. It has been shown that there are internal Shine-Dalgarno sequences in the mRNA that are scanned for by the 3' end of the 16S rRNA during translation, and that pairing between the mRNA and the 16S rRNA at these sites leads to efficient ribosomal frameshifting at these frameshift sites (Weiss *et al.*, 1987). Another location of the 16S rRNA has been implicated in the specific recognition of the UGA termination codon (Murgola *et al.*, 1988). Mutations in this site create a codon-specific translational suppressor that reads through the UGA stop codon, but not the UAG or UAA stop codons.

Subunit association is also mediated by elements in the 16S rRNA. Mutations at the highly conserved nucleotides G791 and A792 cause a dramatic loss in the ability of the ribosome to initiate translation by inhibiting association of the 50S and 30S subunits (Tapprich *et al.*, 1989; Santer *et al.*, 1990). Chemical modification interference using kethoxal, which modifies the Watson-Crick face of guanosine residues, also pinpointed residues in this region as critical for subunit association (Herr *et al.*, 1979). Similar studies have implicated regions in the 3' minor domain of the 16S rRNA as important for subunit association along with the 790 loop (Zweib *et al.*, 1986).

The ribosome contains binding sites for two tRNAs; one site is referred to as the P (peptidyl) site, and the other the A (aminoacyl) site. The tRNA is situated in the ribosome such that the anticodon loop is placed near the decoding site of the ribosome while the aminoacyl arm is located near the peptidyl transferase center of the 50S subunit. The 3'

minor domain of 16S rRNA has been shown to be the decoding site where mRNA-tRNA codon-anticodon base pairing occurs. Both the anticodon of tRNA (Prince *et al.*, 1982; Sylvers *et al.*, 1992) and mRNA (Vladimirov *et al.*, 1990; Juzumiene *et al.*, 1995) show extensive UV induced crosslinks to this region of the 16S rRNA. Another set of studies used chemical probes to detect nucleotides that were shielded from attack by binding of tRNA to the A and P sites (Moazed & Noller, 1986; Moazed & Noller, 1990). These same nucleotides are also protected from chemical modification upon binding of antibiotics that specifically interact with the A and P sites (Moazed & Noller, 1987). Along with highly conserved nucleotides in the 3' minor domain, tRNA- and decoding site-specific antibiotics also protect the 530 loop of 16S rRNA. This highly conserved structure is implicated not only in the decoding site but also in regulating the translational fidelity of the ribosome by proofreading the aminoacyl-tRNAs entering the A site. Certain mutations in this loop abolish the interaction of the EF-Tu-tRNA-GTP complex with the ribosome (Powers & Noller, 1990) while others increase the translational error rate by affecting tRNA selection (O'Conner *et al.*, 1992). These mutations have been shown to disrupt the ability of the 16S rRNA 530 loop to form a pseudoknot structure (Van Ryk & Dahlberg, 1995).

The functional importance of the 16S rRNA is mirrored in the 23S rRNA. Critical functions in translation are played by 23S rRNA in the peptidyl transferase center and recruitment of elongation factors such as EF-Tu and EF-G (Noller, 1993). Despite the dominant role 16S rRNA plays in the function of the 30S subunit, this functionality is only achieved in concert with a number of proteins that specifically bind to this RNA.

Protein-RNA interactions in the 30S subunit

In vitro assembly of the prokaryotic 30S subunit

All of the necessary information to direct the assembly of an intact, functional 30S subunit is encoded within the 16S rRNA and the 21 ribosomal (r-) proteins. This was demonstrated by Nomura and coworkers by reconstituting this subunit from purified 16S rRNA and a mixture of 30S proteins (Traub & Nomura, 1968) and subsequently, using the individually purified components (Held *et al.*, 1973). Using the ability to reconstitute 30S subunits that were functionally active in *in vitro* translation assays, the pathway by which this subunit assembles *in vitro* was determined (Mizushima & Nomura, 1970; Held *et al.*, 1974). The 30S subunit assembly map (Figure 1.6) shows that this process is sequential and highly cooperative. Only six ribosomal proteins (S4, S7, S8, S15, S17, and S20), termed primary binding proteins, are capable of interacting with the nascent 16S rRNA in the absence of the other 30S ribosomal proteins. Other proteins, the secondary and tertiary binding proteins, require the presence of one or more proteins bound to the assembling subunit in order to be stably incorporated into the growing assemblage.

The mechanisms of this cooperative assembly are still poorly understood. There are three plausible models that could possibly account for the dependence of the secondary and tertiary binding proteins on previous binding events (Noller *et al.*, 1989). The first is that the early proteins bind specifically to the 16S rRNA, and the late proteins interact with the assembly via protein-protein interactions. Another possible mechanism is that the early binding proteins bind to the 16S rRNA, which causes conformational changes in the RNA such that binding sites for subsequent proteins are created. Finally, both such mechanisms could be occurring during the assembly of the 30S subunit, and preliminary evidence supports this model.

Studies on the assembly of the 30S subunit elucidated several other properties of this ribonucleoprotein particle. Single-protein-omission experiments were performed in which the subunit was reconstituted lacking a single protein to test the effect each protein had on the assembly process and activity of the subunit (Held *et al.*, 1973). The omission

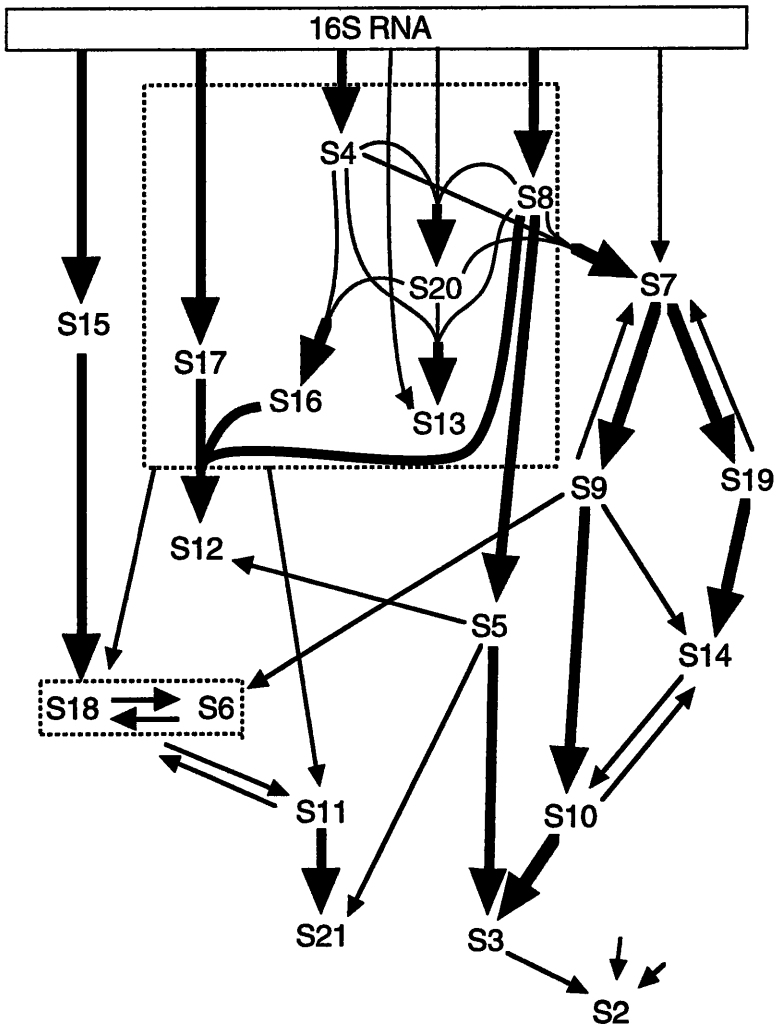


Figure 1.6: The *in vitro* assembly pathway for the *Escherichia coli* 30S subunit. Arrows denote proteins that facilitate the incorporation of other proteins; thick arrows indicate a strong effect. The dashed box around the proteins S6 and S18 indicate that they bind to the assembling subunit as a heterodimer. The arrow from S11 from the large dashed box indicates that the binding of S11 depends upon some of the proteins within the box, although it is unclear exactly which proteins are necessary. Adapted from Held *et al.*, 1974.

of only a few proteins (S1, S6, and S16) had no affect upon the activity of the assembled 30S subunit. Thus, almost all of the ribosomal proteins are required to maintain full functional activity. However, the omission of no single ribosomal protein completely knocked out *in vitro* translation activity. In the original assembly pathway study (Mizushima & Nomura, 1970), one of the primary binding proteins, S15, was not used. Although the omission of this protein might be expected to significantly alter the assembly pathway, when the studies were performed with S15, the overall assembly map changed very little (Held *et al.*, 1974). Instead, it was found that the deficiency in S15 was overcome by a group of other ribosomal proteins that were capable of recruiting the S6/S18 heterodimer to the assemblage (ribosomal protein S1 was not included in the assembly map because of its weak association with the 30S subunit).

It was also determined that functional hybrid 30S ribosomal subunits could be prepared by mixing the RNA and protein from different bacterial species (Nomura *et al.*, 1968). For example, the combination of *E. coli* 16S rRNA and *Bacillus stearothermophilus* 30S r-proteins or *B. stearothermophilus* 16S rRNA and *E. coli* 30S r-proteins both yield subunits that are active in *in vitro* translation assays. Therefore, it appears that the fundamental interactions that govern the assembly of a fully functional 30S subunit have been conserved in prokaryotic phylogeny. A subsequent study of the binding of the *E. coli* primary binding proteins to 16S rRNA showed that *B. stearothermophilus* 16S rRNA is an effective competitor of the *E. coli* 16S rRNA interaction with S4, S7, and S20, further supporting the idea that the protein-RNA interactions in the ribosomes of these two species are very similar (Garrett *et al.*, 1971).

RNA-protein interactions within the ribosome

Since the 30S subunit can be readily reconstituted from its purified constituents, this is an attractive ribonucleoprotein particle to study in detail using biochemical

techniques. Of particular interest is the nature of the many potential protein-RNA interactions. By stepwise reconstitution of the 30S subunit following the assembly pathway of Nomura, the regions of the 16S rRNA that are contacted by the various 30S r-proteins were determined. The earliest studies involved the association of a primary binding protein to 16S rRNA, cleaving the 16S rRNA with ribonucleases, isolating stable protein-RNA complexes by sucrose gradient centrifugation, and sequencing the bound RNA fragments (Zimmerman *et al.*, 1971; Zimmerman *et al.*, 1975; Müller *et al.*, 1979; Zimmerman & Singh-Bergmann, 1979). Using this technique, it was determined that S4 and S20 bind to the 5' domain of 16S rRNA, S8 and S15 bind to the central domain, and S7 binds to the 3' domain. Furthermore, these types of experiments provided support for the assembly map of the 30S subunit. The proteins S6 and S18, whose binding is strongly potentiated by S15, are shown to bind to the same region of RNA as S15 (Zimmerman *et al.*, 1975). Cooperative assembly mediated by local conformational changes in the rRNA caused by S15 binding or direct protein-protein contacts between S6/S18 and S15 would predict that the binding sites of these proteins be adjacent.

More detailed mapping of the protein-RNA interactions within the 30S subunit was accomplished by monitoring the protection of the 16S rRNA by ribosomal proteins from chemical and nuclease probes as the subunit was assembled stepwise (Stern *et al.*, 1986; Powers *et al.*, 1988a; Powers *et al.*, 1988b; Stern *et al.*, 1988a; Stern *et al.*, 1988b; Svensson *et al.*, 1988). Changes in the pattern of chemical modification and nuclease cleavage of the 16S rRNA upon addition of each ribosomal protein to the assembling subunit was monitored by primer extension. The result of these studies showed some striking features. First, all of the ribosomal proteins (except for S1, which was not tested since its interaction with the 30S subunit is not very tight) have an effect on the sensitivity of 16S rRNA to modification upon binding. Furthermore, the primary and secondary proteins tend to show more extensive patterns of chemical modification protections and enhancements upon binding 16S rRNA than the tertiary proteins. This suggests that the

primary and secondary binding proteins bind to the assembling subunit predominantly via protein-RNA interactions, while the tertiary binding proteins interact primarily via protein-protein interactions (Noller *et al.*, 1989). Second, proteins that are linked together in the assembly map footprint to neighboring regions of the 16S rRNA. For instance the ribosomal proteins S15, S6, and S18 all footprint to the central domain, whereas S7, S9, and S19 all footprint to the 3' major domain. This indicates that the primary and secondary binding proteins provide points of nucleation for further protein binding during early assembly (Noller & Nomura, 1987).

These studies suffer from several problems that prevent a detailed picture of the sites of interaction between individual ribosomal proteins and the 16S rRNA from emerging (Powers & Noller, 1995). The chemical probes employed in these studies monitor only those bases that are not involved in pairing interactions with their Watson-Crick face. Thus, interactions with the double-stranded regions of the 16S rRNA and the sugar-phosphate backbone were not explicitly probed. Another problem is that the chemical footprinting data consist of a complex set of enhancements and protections due to protein binding. It was not readily apparent whether a given change in the chemical modification pattern was due to a protein-RNA contact or a conformational change in the RNA itself. Also, many proteins display overlapping sets of modification enhancements and protections, which further complicated the analysis. These problems were mostly overcome in a subsequent study in which hydroxyl radicals generated by Fe^{2+}-EDTA were used to probe the protein-RNA interactions during stepwise assembly of the 30S subunit (Powers & Noller, 1995). This reagent attacks the ribose sugar to generate a site of strand scission, which is detectable as a reverse transcriptase stop. Using this chemical reagent, all nucleotide positions within the 16S rRNA were capable of being probed, and the results of this study were consistent with those using kethoxal and dimethyl sulfate. The primary and secondary binding proteins show extensive regions of protection of the RNA, indicating these proteins form intimate contacts with the 16S rRNA. On the other hand,

many of the tertiary binding proteins display only a few protections of the RNA, suggesting that they are incorporated via mostly protein-protein interactions. Also, the problem of overlapping chemical and nuclease footprints was virtually eliminated; all observed protections of the 16S rRNA are due to the influence of a single ribosomal protein. For instance, the ribosomal protein pairs S15 and S6/S18, and S6/S18 and S11 all show overlapping base protections in the central domain of 16S rRNA using dimethyl sulfate and kethoxal footprinting, but the binding sites of each protein were resolved using the hydroxyl radical probing.

Towards the structure of the 30S subunit

Using the wealth of biochemical data generated for the 16S rRNA, several groups have developed low resolution models of the 30S subunit (Brimacombe *et al.*, 1988; Stern *et al.*, 1988c; Hubbard & Hearst, 1991; Malhotra & Harvey, 1994). In this approach, biochemical data from various experimental techniques were utilized to provide crude internucleotide, interprotein, and RNA-protein distance constraints throughout the 30S subunit, and these data were used in model building programs that utilize similar methods to programs used for refining X-ray crystallographic or NMR data (Malhotra & Harvey, 1994). Interprotein distances were provided by low angle neutron diffraction data (Capel *et al.*, 1987), which yield the three-dimensional positions of the centers of mass of twenty of the small subunit proteins. Distance constraints between elements in the 16S rRNA were provided by numerous studies involving crosslinking agents of different functionalities and lengths (reviewed in Brimacombe, 1995). Sites of RNA-protein interactions were determined from crosslinking data as well as chemical footprinting studies. The global morphology of the 30S subunit has been revealed using various electron microscopic techniques (Brimacombe, 1995).

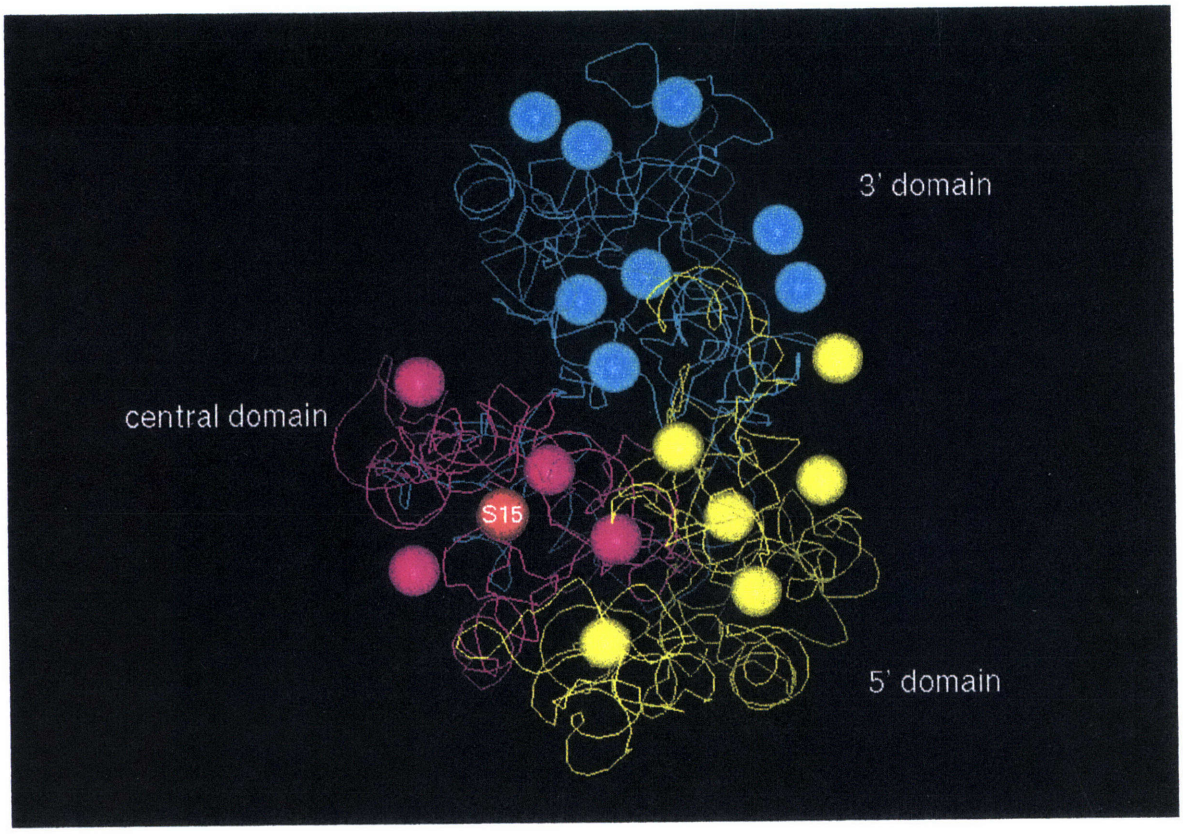


Figure 1.7: Model of the *Escherichia coli* 30S ribosomal subunit developed by Malhotra and Harvey. The 16S rRNA is shown as a ribbon, and the ribosomal proteins are shown as spheres. The 5' domain of the 16S rRNA (nucleotides 1-520) and the associated proteins (S4, S17, S12, S5, S16, and S20) are colored yellow, the central domain (nucleotides 521 -920) and associated proteins (S8, S15, S6, S18, S11, and S21) are displayed in magenta, and the 3' major and minor domains (nucleotides 921-1542) and associated proteins (S7, S9, S19, S14, S13, S10, S2, and S3) are colored blue. Ribosomal protein S1 has been omitted from this figure. Ribosomal protein S15, the focus of studies contained in this thesis, is shown in red.

All of these models (Brimacombe *et al.*, 1988; Stern *et al.*, 1988c; Hubbard & Hearst, 1991; Malhotra & Harvey, 1994) agree quite well with respect to the gross features of the structure and organization of the proteins and 16S rRNA in the 30S subunit, but they lack any reliable information on the local details of this organization. In each model (Figure 1.7 shows a view of the model developed using quantitative model building techniques by Malhotra and Harvey) the 30S subunit is clearly divided into three distinct RNP subdomains, which correlate to the three subdomains of the 16S rRNA. Furthermore, proteins which interact with one another via the assembly pathway are located close in space to one another and are generally in the same subdomain. The regions of RNA in the central and 3' domains of 16S rRNA that are implicated in the A and P sites of the decoding site are all placed close together in space near the cleft of the 30S subunit. Recent cryo-electron microscopy data of tRNA bound to intact 70S subunits clearly place the anticodons of A and P site bound tRNAs in the same location in the 30S subunit, supporting these models (Agrawal *et al.*, 1996).

Despite further attempts to refine these biochemical models of the 30S subunit, detailed structural information will only be achieved using X-ray crystallography and NMR spectroscopy. An ongoing effort has been made to produce diffraction quality crystals of the ribosome and its subunits. Recently, crystals of the 30S subunit that diffract to 4.5 Å have been grown (Glotz *et al.*, 1987), but data on these crystals has been hard to collect because they are extremely sensitive to radiation. Furthermore, generating useful derivatives for solving the phase problem have not been straightforward (Yonath & Franceschi, 1993). Even if a complete medium or high resolution data set can be collected, fitting 21 proteins and 1542 nucleotides of RNA into the electron density map remains a formidable task. Another approach to obtaining detailed structural data on the 30S ribosomal subunit has been to solve the structure of individual proteins. Currently, the structure of only four small subunit proteins has been solved by X-ray and NMR techniques: S5 (Ramakrishnan & White, 1992), S6 (Lindahl *et al.*, 1994), S8 (Davies *et*

al., 1996), and S17 (Jaishree *et al.*, 1996). In part, this is due to the fact that the ribosomal proteins are notoriously difficult to crystallize, and thus despite their ready availability in bulk quantities, the effort to obtain structural data has not been easy. A more current approach that has been taken is to isolate minimal protein-RNA complexes which might be amenable to structural analysis. This approach has been taken to develop minimal RNA binding sites for the primary binding proteins S7 (Dragon & Brakier-Gingras, 1993; Dragon *et al.*, 1994) and S8 (Mougel *et al.*, 1993; Wu *et al.*, 1994) in hopes that these proteins will be more amenable to structural analysis bound to their specific rRNA sites. This is the approach that has been employed in the work presented in this thesis in an effort to develop a minimal complex between the 30S primary binding protein S15 and its specific 16S rRNA binding site.

The ribosomal protein S15

The *Escherichia coli* ribosomal protein S15 is a highly basic protein comprised of 87 amino acids (Morinaga *et al.*, 1976). Currently, the sequences of S15 variants from a number of prokaryotes, archae, and chloroplasts are known (Figure 1.8). Most of the conserved residues and the basic residues are concentrated towards the carboxyl terminal half of the protein. Since basic residues are typically important in protein-nucleic acid interactions, this may indicate that this region of the protein contains sites that interact with the 16S rRNA. The sequence of S15, although highly conserved amongst its own homologs, shows no strong homology to any other ribosomal protein. The only other protein that displays any homology to S15 is human SRP-19, a protein that binds directly to the mammalian SRP RNA (Zwieb, 1992).

Using S15 prepared under denaturing conditions from *E. coli* 30S subunits, some physical studies on this protein suggest that this protein has a discrete tertiary fold with a

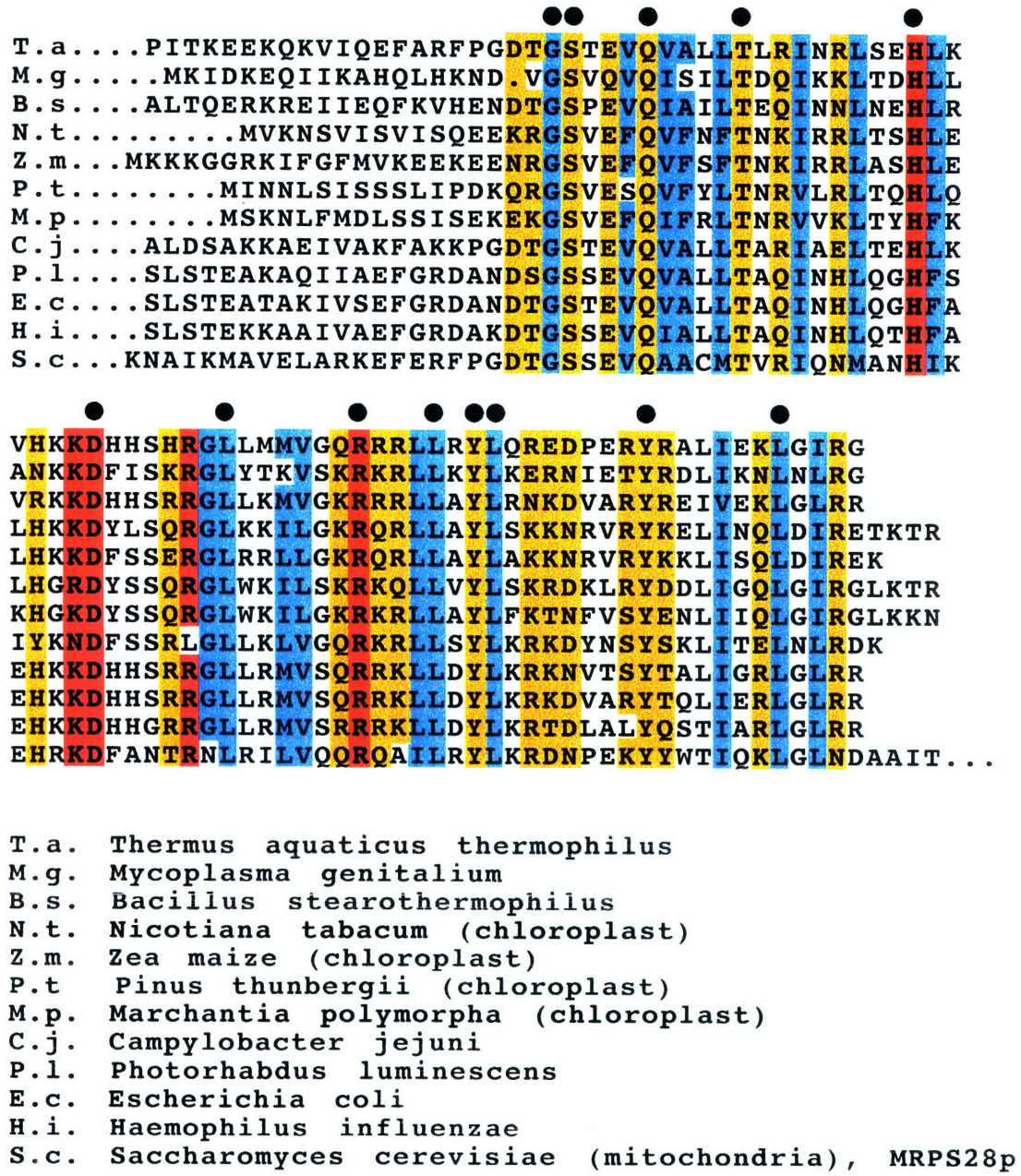


Figure 1.8: Sequences of ribosomal protein S15 from eubacteria and chloroplasts, along with a highly homologous variant from the *Saccharomyces cerevisiae* mitochondrial ribosome. Dots indicate amino acids that are conserved in all of the listed variants. Red shading indicates conserved charged amino acids, orange indicates conserved polar amino acids, and blue indicates conserved non-polar amino acids.

high degree of alpha-helical character . The circular dichroism (CD) spectrum of this protein shows strong minima at 221 and 208 nm, which are characteristic of alpha helices (Gogia *et al.*, 1979). Differences observed in the proton nuclear magnetic resonance (NMR) spectrum of S15 under native and denaturing conditions were interpreted as evidence for tertiary structure (Morrison *et al.*, 1977; Gogia *et al.*, 1979). Microcalorimetric measurements of S15 during heat denaturation demonstrated a distinct heat absorption peak with a temperature of transition around 72 °C (Khechinashvili *et al.*, 1978). This clear peak is indicative of a conformational transition, suggesting cooperative unfolding of the protein, and thus tertiary structure. This behavior of S15 is in marked contrast to many of the other ribosomal proteins, which demonstrate little evidence for extensive secondary or tertiary structure (Morrison *et al.*, 1977).

Interaction with 16S rRNA

The ribosomal protein S15 has been shown to interact specifically with the central domain of the 16S rRNA. Results of various biochemical studies of the binding of S15 to 16S ribosomal RNA and subfragments derived from the central domain are summarized in Figure 1.9. The binding site of S15 as determined by nuclease digestion of a 16S rRNA/S8/S15 complex and by enzymatic and chemical footprinting on 16S rRNA or 16S rRNA fragments is restricted to a local region in the central domain. Nuclease digestion experiments localize S15 binding to helix 21 and 22 of 16S rRNA, and some kethoxal protections of guanines by S15 were shown to be located on helix 22 (Zimmerman *et al.*, 1975; Müller *et al.*, 1979; Zimmerman & Singh-Bergmann, 1979). Enzymatic and chemical footprinting experiments of 16S rRNA and a fragment of the central domain of 16S rRNA suggest that S15 interacts with the 3' end of helix 22 and the nucleotide 724-730 region (Mougel *et al.*, 1988; Svensson *et al.*, 1988). This latter region was later determined to contain a phylogenetically conserved tetraloop, suggesting that S15 is a

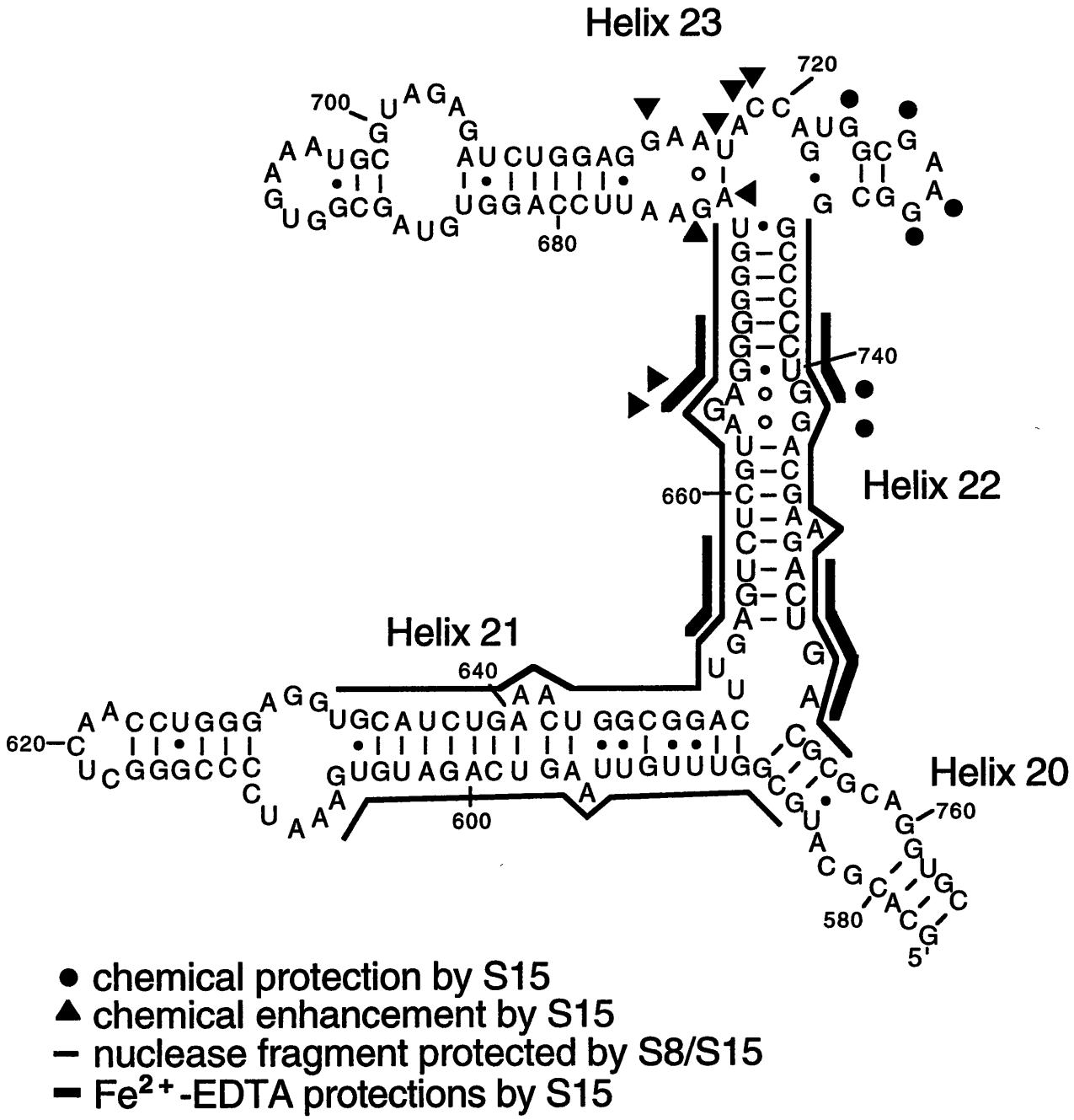


Figure 1.9: Summary of the results of previous studies to localize the S15 binding site in *E. coli* 16S rRNA, superimposed upon the 577-764 region (see Figure 1.2). The biochemical data pertaining to the S15-rRNA interaction are the result of chemical footprinting (Mougel *et al.*, 1988; Svensson *et al.*, 1988), nuclease fragment protection studies (Muller *et al.*, 1979; Zimmerman & Singh-Bergmann, 1979), and Fe^{2+}-EDTA footprinting of S15-16S rRNA complexes (Powers & Noller, 1995).

tetraloop binding protein based upon its homology to SRP-19, which is believed to interact specifically with a GNRA type RNA tetraloop (Zwieb, 1992). However, many of these enhancements and protections of chemical probing overlap extensively with those of the S6/S18 heterodimer, making it hard to interpret which nucleotides are directly recognized by S15. Probing the S15-16S rRNA interaction with Fe^{2+}-EDTA, revealed specific protections mainly in the 5' region of helix 22 centering upon nucleotides in a three-way helical junction and near an internal loop containing several A•G base pairs (Powers & Noller, 1995).

S15 as a repressor of its own translation

Many of the ribosomal proteins fulfill a secondary role as translational repressors (reviewed in Zengel & Lindahl, 1994). Autogenous control of ribosomal protein synthesis is accomplished by a protein that is capable of both acting as a structural component in the ribosome and as a regulatory protein that controls its own expression as well as others in the same operon. These proteins are almost entirely primary binding proteins, capable of directly binding to rRNA and early in ribosomal assembly. One of the implications of this is that autoregulatory control by these proteins is likely to entail recognition of a target in the mRNA that is structurally similar to its binding site in the rRNA. The direct interaction between a number of these translational repressor proteins and their respective mRNA have been demonstrated using footprinting and filter-binding techniques, including S15. For many of these proteins, namely S7 (Nomura *et al.*, 1980), S8 (Cerretti *et al.*, 1988; Gregory *et al.*, 1988), and L1 (Kearny & Nomura, 1987), their binding sites in the mRNA are similar in sequence and secondary structure, providing an obvious explanation for their dual activities. However, for some proteins, like S4 (Deckman & Draper, 1985) and S15 (Philippe *et al.*, 1990), there is no distinct similarity between the mRNA and rRNA binding

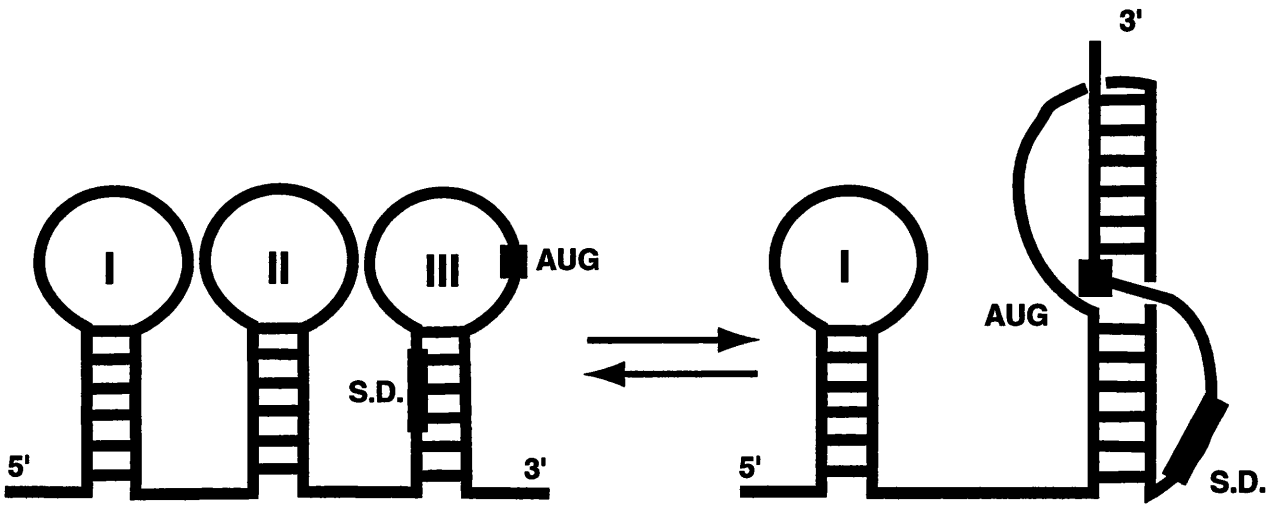


Figure 1.10: Secondary structure representation of the S15 regulatory region of its own mRNA. The S15 repressor site around the ribosome loading site is in dynamic equilibrium between two stem-loop structures (II and III) and a pseudoknot. S15 binds specifically to the pseudoknot form, preventing the formation of a translational preternary complex, thus inhibiting translation of the S15 message. Adapted from Portier *et al* ., (1990).

sites. Thus, how a protein is capable of recognizing two seemingly very different RNA structures poses an interesting twist on the nature of these protein-RNA interactions.

S15 has been shown to regulate its own translation by binding to a region within its mRNA that overlaps with the ribosome loading site (Portier *et al.*, 1990a). This has the effect of inhibiting the formation of a stable preinitiation ternary complex between the mRNA, ribosomal 30S subunit, and $tRNA^{fMet}$ (Phillipe *et al.*, 1993). The mRNA at this site is in dynamic equilibrium between a conformation containing two hairpin loops and a conformation containing a pseudoknot, and S15 preferentially binds to and stabilizes the pseudoknot form (Figure 1.10) (Portier *et al.*, 1990b). Detailed chemical and mutational analysis of the determinants within the pseudoknot that are required for S15 binding show that the protein primarily recognizes a unique conformation of the sugar-phosphate backbone provided by the pseudoknot (Benard *et al.*, 1994; Phillipe *et al.*, 1995). The sequence and structure of this pseudoknot bears little apparent resemblance to any structures in the proposed rRNA binding region, thus providing no additional insight into the RNA elements that contribute to specific recognition by S15.

In order to understand the S15-rRNA interaction on a structural level, whether by X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy, it was imperative to create a minimal protein-RNA complex. Since there was not sufficient information to readily develop a minimal S15-rRNA complex, it was necessary to develop a detailed understanding of this interaction on a biochemical level before proceeding to structural studies.

Chapter 2: Methods for analysis of RNA structure and RNA-protein interactions.

RNA sequencing and secondary structure mapping

Nuclease sequencing and mapping

The sequence and secondary structure of ^{32}P end-labeled RNA transcripts can easily be determined by via limited strand scission catalyzed by base-specific ribonucleases (RNases) (Donis-Keller *et al.*, 1977). Most RNases are robust enzymes, capable of catalyzing strand scission under conditions which denature RNA (7 M urea and elevated temperatures, typically 50 °C) for sequencing applications, as well as in a variety of native conditions that may vary widely in pH and ionic strength in order to map RNA secondary structure (Knapp, 1989). A number of RNases commonly used for sequencing and structure probing are listed in Table 2.1, along with some of their properties.

The most common enzymatic sequencing and structure probe used in these studies was RNase T_1. This enzyme specifically cleaves phosphodiester bonds in the RNA ribose-phosphate backbone 3' to unpaired guanosines, to generate fragments with a 2'/3' phosphate and a 5' hydroxyl group (Ehresmann *et al.*, 1987). Sequencing under denaturing conditions was performed in 20 µl volumes using approximately 1 pmol of end-labeled RNA (5' or 3') (see the Appendix 1 for RNA labeling methodology) and 1 µg of *E. coli* tRNA (United States Biochemical) in a buffer containing 20 mM sodium citrate, pH 5.0, 1 mM EDTA, 7 M urea, and 0.025% (w/v) xylene cyanol (Knapp, 1989). This reaction mixture was preincubated at 50 °C for five minutes prior to the addition of 1 unit

Table 2.1: Enzymes and chemicals used for RNA sequencing, determining secondary structure, and probing S15-RNA interactions.

RNA probe	specificity	molecular weight	detection†
RNase T_1	unpaired GpN	11,000	either
RNase U_2	unpaired ApN>GpN	12,490	either
RNase CL_3	unpaired CpN	16,800	either
RNase *B. cereus*	unpaired UpN or CpN	N/A	either
RNase Phy M	unpaired UpN or ApN	N/A	either
RNase S_1	unpaired N	32,000	either
RNase V_1	ds RNA	15,900	either
DMS	G-N7	126	either*
	A-N1		reverse tx
	C-N3		reverse tx
DEPC	A-N7>G-N7	174	either*
hydrazine(aq)	uracil	50	either*
hydrazine/NaCl	cytosine	50	either*
ENU	phosphates	117	either*
CMCT	G-N1	424	reverse tx
	U-N3		reverse tx
iodine	phosphorothioate	254	either

† Detection of the site of modification in the RNA can be detected using: (either), end-labeled RNA or by reverse transcription (an asterisk indicates that a separate chemical cleavage step in necessary prior to detection); (reverse tx) this modification can be detected by reverse transcription of the RNA only. Adapted from Ehresmann *et al.*, 1987.

T_1 nuclease (United States Biochemical). The length of incubation varied according to the length of RNA probed and its guanosine content. For an RNA of around 200 nucleotides in length, this reaction was allowed to proceed for 14 minutes. For shorter RNAs, the incubation time necessary for the proper level of cleavage increased; for RNAs of 60-90 nucleotides in length, the RNAs were incubated for 25-30 minutes. The proper level of cleavage for a particular RNA was determined by incubating the RNA for various times with 1 unit of RNase. After incubation, the reactions were immediately analyzed by loading 6 µl onto an 8-20% (depending on the length of RNA being sequenced) denaturing 29:1 acrylamide:bisacrylamide gel and visualized using autoradiography (Sambrook *et al.*, 1989). The proper level of RNA modification that yields reliable sequence information is statistically one hit per molecule. Thus, to ensure that the RNA is not overmodified, the reaction time was set such that no less than 50% of the total RNA remained uncleaved. The T_1 sequencing reaction was always accompanied by an alkaline hydrolysis ladder, in which hydroxide was used to uniformly cleave the RNA backbone at all phosphate positions, again at a statistical level of one hit per molecule. Uniform hydrolysis of RNA is performed in a 10 µl reaction containing 1 pmol of end-labeled RNA in Alkaline hydrolysis buffer (50 mM $NaHCO_3/Na_2CO_3$, pH 9.2, 1 mM EDTA) and incubated for five minutes at 90 °C. The reaction was quenched by addition of 10 µl Urea stop mix (9.8 M urea, 1.5 mM EDTA, 0.025% xylene cyanol). By running this reaction alongside sequencing lanes, the nucleotide spacing between bands in the RNase T_1 sequencing lane can be determined, thus establishing the sequence identity of each guanosine.

Since RNase T_1 only cleaves at unpaired guanosine residues, by subjecting an RNA to T_1 catalyzed strand scission under conditions in which it is stably folded, regions of single- and double-strandedness can be determined. Like RNA sequencing, the reaction conditions that yield the appropriate level of cleavage of native RNA were determined empirically by monitoring the cleavage as a function of time for a given amount of enzyme. For structure mapping, 1 pmol of end-labeled RNA (5' or 3') and 1 µg of *E. coli* tRNA

were incubated in a 20 µl reaction with a buffer containing 10 mM K-HEPES, pH 7.5, 50 mM potassium acetate, 1 mM EDTA with 1 unit of RNase T_1 at 37 °C. For a 200 nucleotide RNA, this reaction was allowed to proceed for fifteen minutes, quenched with 20 µl of Urea stop mix, and analyzed using polyacrylamide gel electrophoresis. Guanosine residues that are in unstructured regions demonstrate substantially higher levels of cleavage over guanosine residues in structured regions of the RNA when this reaction is compared to a denaturing sequencing ladder.

Other RNases in conjunction with RNase T_1 can be used to provide more detailed sequence and structural information. Complete RNA sequence information can be obtained by sequencing end-labeled RNA with a combination of RNase T_1, RNase U_2 (adenosine specific), RNase CL_3 (cytidine specific), RNase PhyM (adenosine and uridine specific), and RNase, *B. cereus* (cytidine and uridine specific) (Table 2.1). Since all of these enzymes cleave RNA to yield products with a 2'/3' phosphate and a 5' hydroxyl group, the cleavage products of all of these enzymes can be aligned with an alkaline hydrolysis ladder, allowing for unambiguous sequencing. Although these enzymes should yield uniform cleavage throughout the RNA under denaturing conditions, this was not always the case; RNase CL_3 and RNase *B. cereus* often did not yield uniform cleavage ladders, which seriously complicated sequencing. Thus despite the technical ease of RNase sequencing, chemical sequencing proved to be far more reproducible and reliable.

Secondary structure mapping by RNase T_1 is often accompanied by RNase S_1 (a single-stranded specific endonuclease) and RNase V_1 (a double-stranded specific endonuclease) to fully probe the RNA. While easy to use, these nucleases suffer from problems that limit their applicability. Nuclease S_1 behaves well at its optimal pH (4.5), but for structure mapping under neutral conditions much higher concentrations of S_1 are required (Ehresmann *et al.*, 1987). S_1 requires multiple single-stranded nucleotides to efficiently cleave RNA (Wurst *et al.*, 1978), and thus single nucleotide bulges and internal loops are not revealed by this enzyme. Furthermore, the cleavage of unpaired nucleotides

can be blocked by the folded conformation of the RNA, and thus, not all single-stranded regions of the RNA will be revealed using any of the single-strand specific RNases, since all of these enzymes suffer from the problem of being sterically bulky. Similarly, RNase V_1 will not reveal all of the nucleotides that are within duplex regions of an RNA. Although RNase V_1 specifically cleaves nucleotides in double stranded helices, it requires at least five contiguous base pairs, and thus the cleavage efficiency of this enzyme varies considerably within duplex regions (Wyatt & Walker, 1989). Furthermore, this enzyme tends to cleave a duplex region asymmetrically, cleaving one side of the duplex more strongly than the other (Puglisi, 1989). Therefore, while RNases have been extensively utilized to probe RNA structure and protein-RNA interactions, these problems severely limit their capability to generate high resolution biochemical data.

Chemical sequencing

The alternative technique for RNA sequencing is to use chemicals to cleave RNA according to nucleotide type (Peattie, 1979). This is analogous to the procedure developed to sequence DNA directly by Maxam and Gilbert (Maxam & Gilbert, 1977), although the chemistry differs due to the differences in the chemical properties of RNA and DNA. Like enzymatic sequencing, this requires RNA that is radioactively end-labeled, however, because of the chemistry of strand scission employed, it must labeled at the 3' end with [5'-^{32}P]pCp (England & Uhlenbeck, 1978). In this technique, the sequence is read from four sequencing lanes corresponding to: guanosine (G), adenosine (A>G), cytidine (C>U), and uridine (U). For chemical sequencing, the modification reactions are performed under denaturing conditions such that the all bases are modified with equal frequency. Under native conditions, dimethyl sulfate and diethyl pyrocarbonate preferentially modify bases that are in single stranded, unstructured regions (Peattie & Gilbert, 1980), and thus could be used to map the RNA secondary structure.

Guanosine sequencing is performed using the chemical reagent dimethylsulfate (DMS), which methylates the N7 position, placing a positive charge on this nitrogen (Brookes & Lawley, 1961; Lawley & Brookes, 1963) as shown in Figure 2.1. Although this modification is stable at neutral pH, the 7,8 double bond is easily reduced using sodium borohydride (Wintermeyer & Zachau, 1975), making the glycosidic bond between the guanine base and the ribose sugar subject to hydrolysis under acidic conditions. This allows for subsequent aniline-mediated hydrolysis of the phosphate backbone at the newly formed abasic site. This chemical reaction was performed essentially as described in Peattie and Gilbert (Peattie, 1979). Approximately 50,000 counts per minute (cpm) of 3' end-labeled RNA (in 2 to 5 µl) and 1 µl of 10 mg/ml tRNA was added to 300 µl of DMS reaction buffer (50 mM sodium cacodylate, pH 5.5, 1 mM EDTA), and chilled on ice. The tRNA used for all chemical reactions was extracted twice with 25:24:1 phenol:chloroform:isoamyl alcohol and twice with 24:1 chloroform:isoamyl alcohol to remove any protein impurities. 1 µl of DMS (Sigma) was added to the reaction, quickly vortexed, and placed in a 90 °C waterbath for 1 minute. As with all sequencing reactions, the amount of DMS added and the time of incubation was adjusted such that the RNA is statistically modified with one modification per molecule. The reaction was quenched by the addition of 75 µl ice-cold DMS quench buffer (1.0 M Tris-acetate, pH 7.5, 1.0 M 2-mercaptoethanol, 1.5 M sodium acetate) and 900 µl 100% ethanol. The RNA was precipitated by incubating the reaction for 10 minutes in a dry ice/ethanol bath and pelleted by centrifugation for 30 minutes at 4 °C. The RNA was resuspended in 200 µl 0.3 M sodium acetate, pH 5.2 and reprecipitated with 800 µl 100% ethanol, incubated in a dry ice/ethanol bath, and pelleted. This pellet was washed with 70% ethanol to remove any residual salt and DMS, and dried for 2 to 5 minutes in a Speed-Vac (Savant). The 7,8 double bond was reduced by resuspending the pellet in 10 µl 1.0 M Tris-Cl, pH 8.2, and

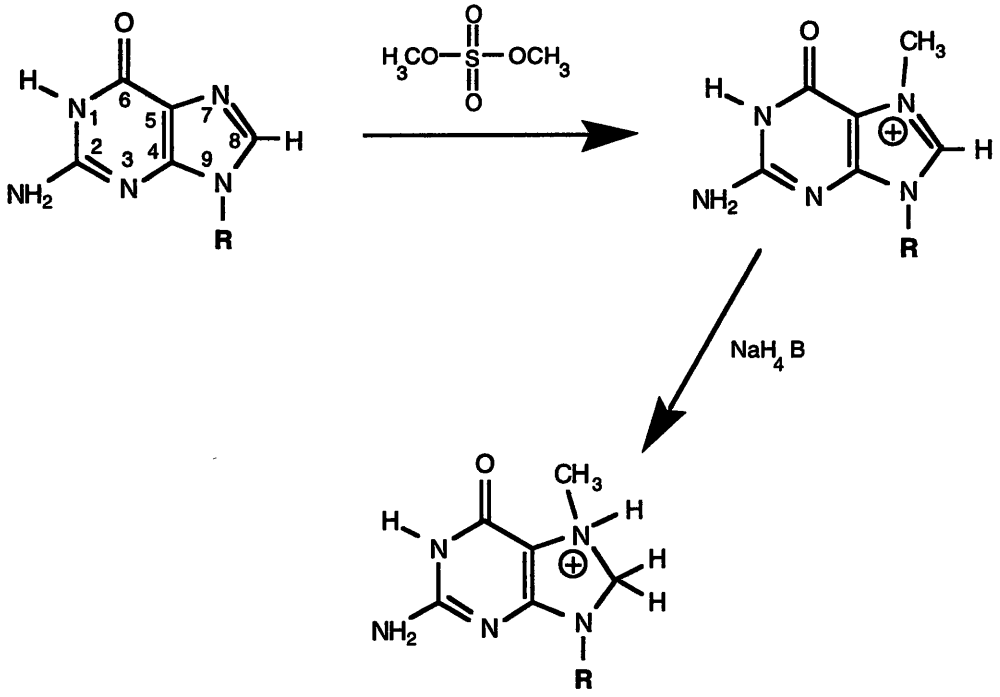


Figure 2.1: Methylation of the guanine N7 position by dimethyl sulfate.

10 µl of fresh 0.2 M sodium borohydride. This reaction was incubated on ice in the dark for 30 minutes, quenched by addition of 200 µl of a buffer containing 0.6 M sodium acetate, 0.6 M acetic acid, pH 4.5, and 0.025 mg/ml tRNA, and precipitated by addition of 600 µl 100% ethanol. After pelleting, washing with 70% ethanol, and drying, the modified RNA was ready for strand scission with aniline.

Diethyl pyrocarbonate (DEPC) reacts with the N7 position of purine bases in aqueous solution to carboethoxylate this position (Leonard *et al.*, 1971; Vincze *et al.*, 1973), leading to the opening of the imidazole ring by water (Figure 2.2). Under the conditions employed for RNA sequencing, the adenine base reacts 5 to 7 times faster than the guanine base (Ehrenberg *et al.*, 1976), and thus in the sequencing ladder these two bases can be distinguished by their differing band intensities, along with comparison to a DMS sequencing ladder. The DEPC chemical reaction was performed essentially as described in Peattie and Gilbert (Peattie, 1979), except that the volume of DEPC used was typically higher to ensure proper levels of modification. Approximately 50,000 cpm of 3' end-labeled RNA (in 2 to 5 µl) and 1 µl of 10 mg/ml tRNA was added to 200 µl of DEPC reaction buffer (50 mM sodium acetate, pH 4.5, 1 mM EDTA), and chilled on ice. To this reaction mixture, 3 µl of DEPC was added, incubated at 90 °C for one minute, and quenched by the addition of ice-cold 50 µl 1.5 M sodium acetate and 750 µl 100% ethanol. The RNA was precipitated for 10 minutes in a dry ice/ethanol bath and pelleted. The pelleted RNA was resuspended in 200 µl 0.3 M sodium acetate, pH 5.2 and reprecipitated with 800 µl 100% ethanol, incubated in the dry ice/ethanol bath, and pelleted. After washing the RNA with 70% ethanol, and drying for 2 to 5 minutes in a Speed-Vac, it was ready for aniline strand scission.

The pyrimidine bases of nucleic acids are effectively removed by reaction with hydrazine (Temperli *et al.*, 1964). This occurs by the nucleophilic addition of hydrazine across the pyrimidine 5,6 double bond to create a pyrazolone ring (Figure 2.3) (Cashmore & Peterson, 1978), which makes the glycosidic bond susceptible to cleavage (Peattie,

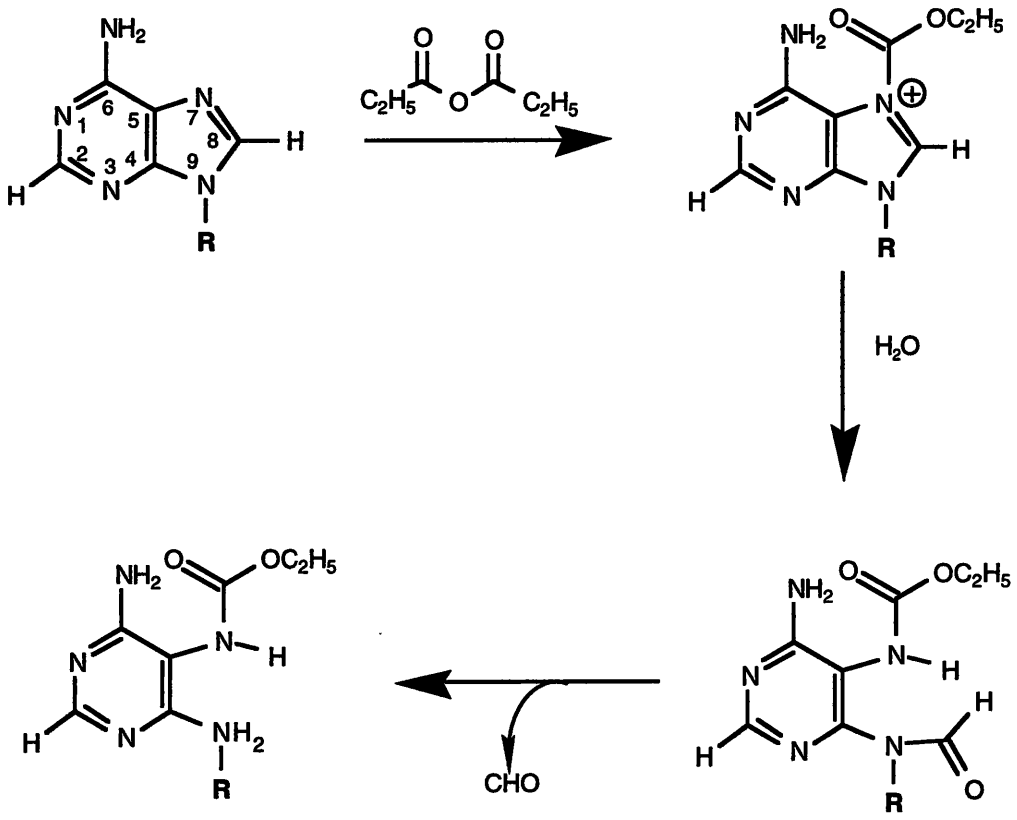


Figure 2.2: Reaction of diethyl pyrocarbonate (DEPC) with the adenine base of RNA.

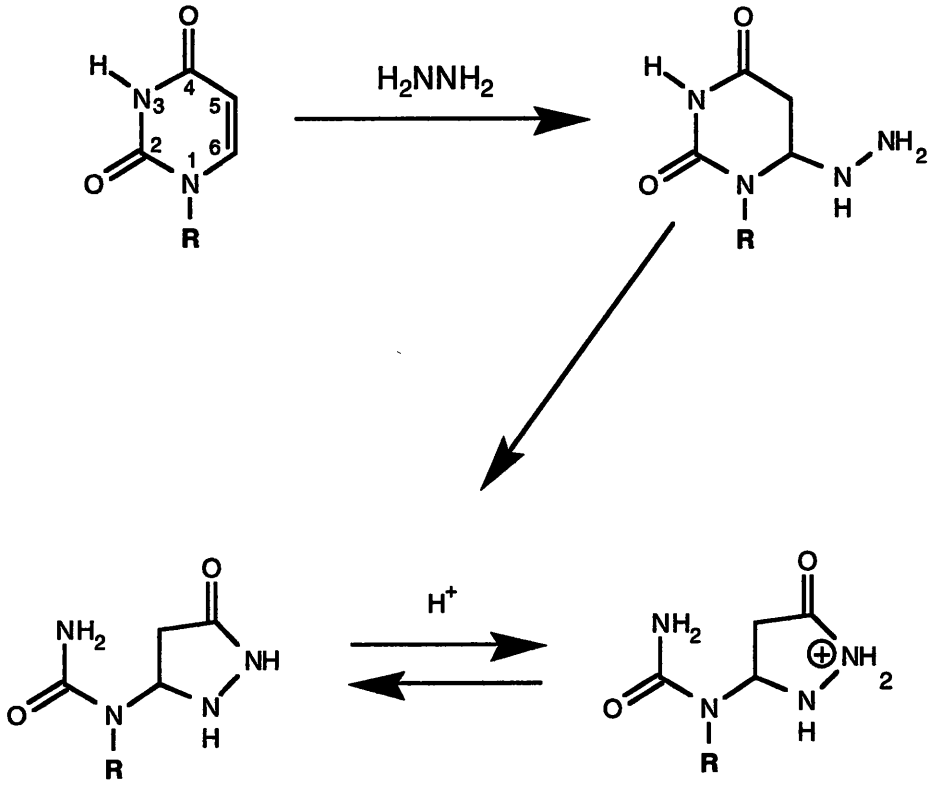


Figure 2.3: Hydrazinolysis of the uracil base in RNA.

1979; Shabarova & Bogdanov, 1994). Under aqueous conditions the level of reactivity is such that the uracil base is much more sensitive to hydrazinolysis than the cytosine base (Verwoerd & Zillig, 1963), whereas in the presence of 3 M sodium chloride under anhydrous conditions, hydrazinolysis of uracil in suppressed and that of cytidine enhanced, as is found in DNA (Maxam & Gilbert, 1977; Peattie, 1979), which allows for uracil- and cytidine-specific RNA sequencing ladders. To perform the uracil-specific reaction, 50,000 cpm of RNA and 10 µg of tRNA were lyophilized to dryness. To this pellet, 10 µl of cold 50% hydrazine/50% water were added, vortexed briefly, and incubated on ice for 10 minutes. The reaction was quenched by addition of 200 µl of cold 0.3 M sodium acetate, pH 5.2, 1 mM EDTA, and 750 µl of 100% ethanol, precipitated for 10 minutes in a dry ice/ethanol bath, and pelleted. The RNA was resuspended in 200 µl 0.3 M sodium acetate, pH 5.2, precipitated by adding 800 µl 100% ethanol, and pelleted. The pellet was washed with 70% ethanol, dried for 5 to 10 minutes in a Speed-Vac, and then subjected to aniline cleavage. The cytosine-specific reaction was performed essentially like the uracil-specific reaction, except that the lyophilized end-labeled RNA/tRNA mixture was resuspended in 10 µl anhydrous hydrazine containing 3 M sodium chloride, vortexed, and incubated for 10 minutes on ice. This reaction was quenched by the addition of 1 ml of 80% ethanol, and subsequently treated the same as the uracil-specific reaction.

Modification of nucleotide bases in RNA by DMS, DEPC, or hydrazine creates an N-glycosidic bond that is sensitive to cleavage by water under mildly acidic conditions which creates an abasic site and exposes the glycosidic hydroxyl group of the ribose sugar (Shabarova & Bogdanov, 1994). This ribose anomeric hydroxyl group undergoes tautomeric exchange with the aldehyde form, which readily reacts with aniline to yield an aldimine (Schiff's base) (Wintermeyer & Zachau, 1970). The α proton is subsequently removed by aniline, resulting in a β-elimination reaction that causes strand scission (Figure 2.4). This reaction is performed under acidic conditions (pH 4.6) which is optimal for stand scission (Turchiniskii *et al.*, 1970), yet mild enough to prevent non-specific

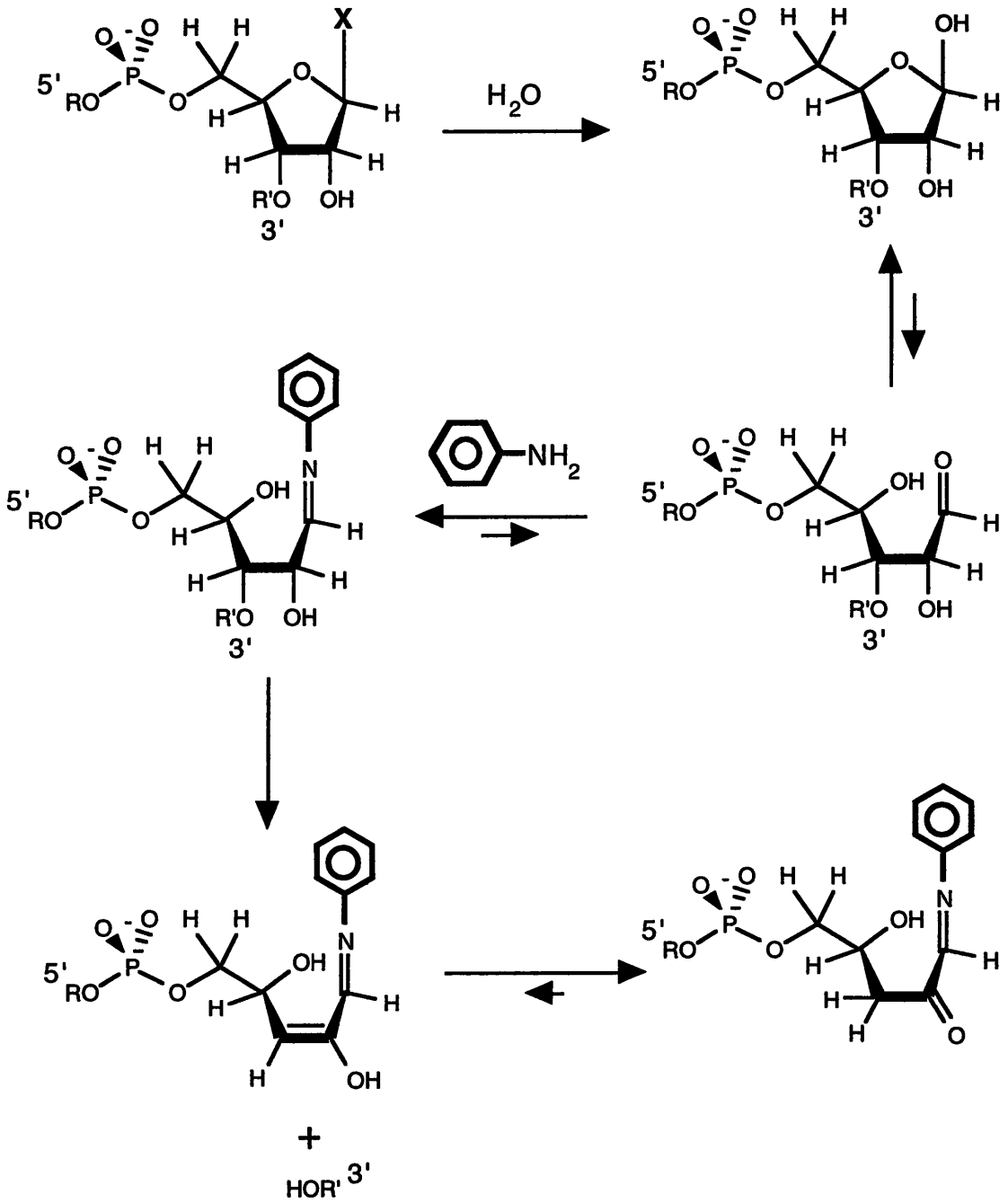


Figure 2.4: Aniline mediated strand scission of chemically modified RNA.

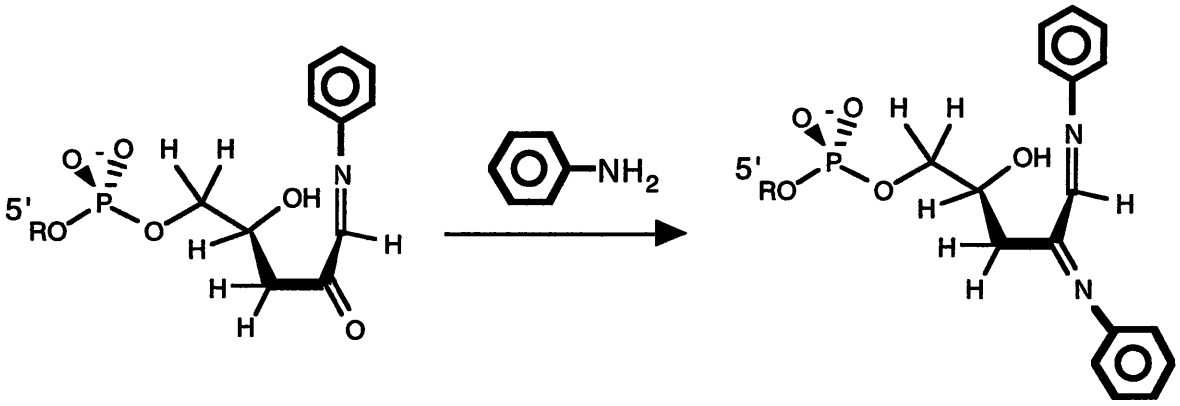


Figure 2.5: An alternative transformation of the 5' product of aniline mediated strand scission. The presence of these alternative products makes using 5' end labeled RNAs in sequence analysis impractical.

hydrolysis of the ribose-phosphate backbone and cleavage of the N-glycosidic bonds of purine bases (Peattie, 1979). A second aniline can react with the ribose sugar of the 5' product of strand scission at the site of cleavage (Figure 2.5) which results in diffuse bands in the RNA sequencing ladder when 5' end-labeled RNA is used. Therefore, for direct chemical sequencing of RNA, 3' end-labeled substrate is always used. Aniline mediated strand scission was performed by dissolving the RNA pellet in 20 µl of freshly made aniline buffer (1 M aniline, 1 M acetic acid, pH 4.5), and incubating in a 60 °C waterbath in the dark for 20 minutes (Peattie, 1979). Prior to use, the aniline (reagent grade) was redistilled under argon to remove chemical impurities that may react with RNA and stored under argon in amber bottles at -20 °C. The strand scission reaction was terminated by freezing in a dry ice/ethanol bath, and the aniline was removed by lyophilizing the reaction to dryness. Residual aniline was removed by resuspending the pellet in 20 µl of ddH_2O, freezing, and lyophilizing twice. The resulting RNA pellet was dissolved in 10 µl ddH_2O and 10 µl Urea stop buffer, and electrophoresed on a denaturing sequencing gel and visualized as described above.

Gel Electrophoresis Mobility Shift Assay

The gel electrophoresis mobility shift assay is a popular and powerful tool for the analysis of protein-nucleic acid interactions. This assay has been successfully applied to the study the interactions between proteins and DNA (Fried & Crothers, 1981; Garner & Revzin, 1981), and proteins and RNA (Hall & Stump, 1992; Weeks & Crothers, 1992; Long & Crothers, 1995), and complex ribonucleoprotein assemblages (Dahlberg *et al.*, 1969). This technique is based upon the observation that the electrophoretic mobility of a nucleic acid in a polyacrylamide gel matrix is altered upon binding proteins and other ligands (reviewed in Carey, 1991). The technical simplicity of this assay makes it a highly attractive method to explore protein-RNA interactions and its ability to provide quantitative

information about these interactions has been firmly established (Carey, 1988; Revzin, 1989; Senear & Brenowitz, 1991). Using variations of this assay, information about equilibrium binding constants, stoichiometry, association and dissociation kinetics, and the salt and pH dependence of protein-nucleic interactions can be obtained.

Before a thorough analysis of a protein-RNA interaction can be undertaken using the gel mobility shift assay, a number of issues must be considered and addressed. The first consideration is the reaction conditions under which the protein and RNA are allowed to bind prior to gel electrophoresis. In these studies, both the protein and RNA were purified to homogeneity (see Appendix 1), so the problems of nuclease contamination or additional protein factors competing for RNA binding, as is often the case for proteins derived from crude cell extracts, did not plague this system. Instead, the major problem was determining the conditions of the binding reaction that yielded a clean and reproducible protein-RNA gel shift. The first consideration was salt concentration. Ribosomal reconstitution and studies of individual ribosomal protein-RNA interactions typically have been performed under high salt conditions; the typical ribosomal reconstitution buffer contains 300 mM potassium chloride and 20 mM magnesium chloride (Traub *et al.*, 1971; Stark *et al.*, 1984). While high ionic strength conditions may be necessary for formation of the complete assembly of the 30S ribosomal subunit, the interaction of any single ribosomal proteins with rRNA may not require such high salt concentrations. Since salt creates artifacts that affect the nucleic acid bands in the gel, it is highly advantageous to minimize the salt concentration present in the reaction mixture. By observing the S15-rRNA gel shift under a number of salt conditions in the reaction mixture, it was empirically determined that this protein-RNA interaction did not require the presence of divalent cations such as magnesium, unlike many other protein-RNA interactions, and that binding occurred in moderate concentrations (25-50 mM) of monovalent cations. Similarly, the pH of the reaction mixture has been shown to have a significant impact upon the observed gel shift (Carey, 1988). Most ribosomal protein-RNA interactions have been studied at pH

7.5, and since this is close to the pH of the Tris-borate running buffer of the gel (pH 8.1), this parameter was not explored for these studies.

A recurrent problem observed in the mobility shift assay is aggregation of the protein-nucleic acid complex. Frequently, upon addition of high concentrations of protein to a nucleic acid probe, an aggregate forms which remains in the loading well of the gel or causes the nucleic acid band to disappear altogether. Without the presence of several components to the binding reaction, concentrations of BS15 greater than 500 nM causes RNA probes to aggregate and remain in the well, a problem we dubbed the "well shift". Addition of unfractionated yeast tRNA to the reaction mixture virtually eliminated this problem. Since tRNA acts as a nonspecific competitor RNA for binding S15, the S15-RNA gel shift was monitored as a function of tRNA concentration. It was empirically determined that 4 μM tRNA was sufficient to fully alleviate the aggregation phenomena, while at the same time minimizing the concentration of nonspecific competitor RNA in the binding reaction (however, this still yielded a ~10,000-fold excess of nonspecific tRNA over specific RNA in the binding reaction). Another problem related to the aggregation phenomenon was variable activity of the protein. Initially, it was observed that the concentration of protein required to shift all of the same RNA probe was variable between experiments and the transition from free RNA to bound RNA was extremely sharp, going from less than 10% bound to over 90% bound in less than a two-fold increase in protein concentration. The addition of the nonionic detergent Nonidet P-40 (Sigma Chemical Co.) to the binding reaction resolved this problem, and resulted in a 10-fold decrease in the apparent equilibrium dissociation constant. This effect of Nonidet P-40 has also been observed for other protein-nucleic acid interactions (Brown *et al.*, 1990; Weeks & Crothers, 1991). Protein aggregation was also diminished by the addition of low concentrations of heparin to the reaction mixtures (typically 5 μg/ml), however, at higher concentrations, heparin is a potent inhibitor of this protein-RNA complex, and thus heparin concentrations were kept as low as possible.

Another important issue considered was the conditions under which the complexes were resolved from free RNA in a polyacrylamide gel. The running buffer employed for these studies, 0.045 M Tris-borate, pH 8.1, 0.001 M EDTA (0.5x TBE), is fairly typical for this assay. For the BS15-rRNA interaction, this buffer sufficed to yield reproducible results and well focused bands. For gel electrophoretic studies of RNA conformation it was necessary to eliminate EDTA in the running buffer in order to add various divalent and polyvalent salts, such as magnesium chloride, hexamine cobalt, and spermidine. The other critical parameter that was varied in the studies of this RNA-protein complex was the acrylamide concentration and crosslinking ratio of the gel matrix. It was observed that in one gel matrix the protein apparently did not shift the RNA upon binding, but changing the gel matrix revealed a pronounced change in mobility. The mobility shift corresponding to S15 interacting with larger RNAs (Fr1 to Fr6 RNA, Figure 3.3) was readily observable in 8% acrylamide gels with a 29:1 acrylamide:bisacrylamide ratio. However, smaller RNAs (Fr7 to Fr16 RNA, Figure 3.3) required much higher acrylamide concentrations (15-20%) and a 19:1 acrylamide:bisacrylamide ratio to visualize the mobility shift. Once a reproducible gel mobility shift has been achieved, a number of experiments can be performed to assess the RNA binding properties of a protein.

Determination of the apparent equilibrium dissociation constant

The first experiment to perform is a measurement of the apparent equilibrium dissociation constant, K_d, of the RNA-protein complex. This is accomplished by titrating a radiolabeled RNA probe with increasing concentrations of protein, and determining the concentration of protein that is required to half-saturate the RNA. In this experiment, it is critical to maintain the concentration of RNA probe such that it is negligible compared to the protein concentration required to half saturate the RNA in order to determine an accurate dissociation constant (Carey, 1991).

Equilibrium dissociation constants were measured between S15 and various RNA species using the following protocol. A constant concentration of labeled probe RNA (approximately 0.01 pmol) was incubated with various concentrations of protein in 10 mM K-HEPES, pH 7.5, 50 mM potassium acetate, 0.1 mM EDTA, 0.1 mg/ml tRNA, 5 μg/ml heparin, and 0.01% Nonidet P-40. RNA to be used in each experiment was heated to 90 °C for 1 minute and cooled on ice for at least 2 minutes prior to adding protein. Reaction volumes of 20 μl were incubated for one hour at 25 °C and one hour at 4 °C before adding 2 μl of Type III Load buffer (30% glycerol, 0.25% bromophenol blue, 0.25% xylene cyanol) (Sambrook *et al.*, 1989). The samples were immediately loaded onto a 8% 29:1 acrylamide:bisacrylamide gel, in 0.5x TBE buffer. For RNA probes Fr1 through Fr6 (Figure 3.3), an 8% 29:1 acrylamide:bisacrylamide gel could effectively resolve the complex; for the other RNA probes, a 20% 19:1 acrylamide:bisacrylamide gel in 0.5x TBE buffer was used. Gels (36 cm wide, 24 cm long, and 0.05 cm thick) were pre-run for one hour at 25 V/cm at 4 °C. Sample volumes of 6 μl were loaded onto the gel with the current on, and run for 4 to 16 hours depending on the RNA probe used.

To quantitate the gel mobility shift assay, the gel was dried and exposed for sixteen hours to a phosphor plate and imaged using a phosphorimager (Molecular Dynamics). For each lane, the bands corresponding to free and bound RNA were boxed and quantitated using ImageQuant software (Molecular Dynamics). At each protein concentration the fraction of bound RNA was calculated using the equation

$$\Theta = \frac{\text{counts}_b}{\text{counts}_f + \text{counts}_b} \qquad (1)$$

where Θ is the fraction of RNA bound, counts_b are the phosphorimager counts in the bound RNA band, and counts_f are the phosphorimager counts in the free band. These data

were fit by nonlinear least squares analysis, using Igor software (version 1.26) for the Macintosh (Wavemetrics). Apparent dissociation constants from direct titration of the RNA probe with BS15 were estimated using the equation

$$\Theta = \frac{[\text{protein}]}{[\text{protein}] + K_d} \qquad (2)$$

where K_d is the apparent dissociation constant.

Determining the stoichiometry of the protein-nucleic acid interaction

All of the ribosomal proteins, except L7 and L12, are present in the ribosome as a single copy. Therefore, an observed protein-RNA complex in a native gel should reflect this 1:1 stoichiometry. The stoichiometric ratio of protein:RNA can be measured by the gel mobility shift assay by slightly modifying the procedure used for equilibrium binding constant measurements. In this experiment, the concentration of the RNA probe must be above the measured K_d by at least five-fold (for S15, which displayed an apparent K_d of 35 nM, an RNA concentration of 282 nM was used). Under these conditions, all of the protein in the binding reaction binds to the RNA until all of the specific protein binding sites in the RNA are occupied, and then no further specific binding occurs. When this data is plotted on a graph of the fraction RNA bound versus the protein:RNA ratio, the fraction of RNA bound should increase linearly until it is saturated, and then plateaus; the break point corresponds to the protein:RNA ratio of the complex.

Determining the relative binding constants of RNAs

One of the problems associated with the gel mobility shift assay is that the magnitude of the shift is dependent upon many factors, including the size and conformation of the RNA, the net charge on the protein at the pH of the running buffer, and the size of

the protein. While direct binding titrations for each RNA could be achieved by changing the gel matrix, it is simpler to use a single RNA that yields an easily observed mobility shift to probe the interaction between the protein and a variety of other RNAs. This can be accomplished using another variation of the gel shift assay in which a complex between the protein and trace concentrations of a labeled probe RNA (Fr1 RNA (Figure 3.3) was used for all of the competition assays in this thesis) is competed away with increasing concentrations of an unlabeled RNA. Typically, sufficient protein is added to the reaction to yield approximately 80% saturation of the labeled RNA probe in the absence of unlabeled competitor. The lower the protein concentration in the competition assay, the lower the concentration of unlabeled competitor RNA is required to compete away half of the complex between the protein and labeled probe RNA. This is important since some of the more weakly binding mutant competitor RNAs required micromolar concentrations to effectively compete.

Competition assays were performed using the standard gel mobility shift assay, using wild-type Fr1 RNA (Figure 3.3) as the labeled probe at a concentration of 0.75 to 1.5 nM, BS15 at a concentration of 50 nM, and varying the concentration of unlabeled competitor RNA. The unlabeled competitor RNA was heat/cooled in TE buffer (10 mM Tris-Cl, pH 8.0, 1 mM EDTA), as described above, before adding to the reaction. The competition assay was quantitated using the same techniques as above to determine the fraction of labeled RNA bound at each concentration of competitor RNA. The competitive binding of protein between the labeled probe RNA and unlabeled competitor RNA is described by the equation

$$\Theta = \frac{P_t(1-\Theta)}{K_T[(1+C_t)K_C] + T_t(1-\Theta)} \qquad (3)$$

where P_t, T_t, C_t, K_T, and K_C are the concentrations of the protein, radiolabeled RNA probe, competitor RNA probe, and the dissociation constants for the wild type and competitor RNA probes, respectively (Lin & Riggs, 1972). This equation can be solved for Θ to yield the physically relevant root

$$\Theta = \frac{1}{2T_t}\{K_T + P_t + (K_T / K_C)C_t + T_t - \sqrt{[K_T + P_t + (K_T / K_C)C_t + T_t]^2 - 4T_tP_t}\} \quad (4)$$

which can be used to fit the data to determine the best value for K_C (Weeks & Crothers, 1992; Long & Crothers, 1995). The apparent relative binding constant, K_{rel}, is the ratio of the apparent dissociation constants of the mutant RNA to the wild type RNA.

Kinetics

It has been recently observed that the gel mobility shift assay is unable to accurately measure subnanomolar peptide-RNA binding affinities, underestimating the apparent dissociation constant by almost 100-fold (Long & Crothers, 1995). This behavior was suggested to reflect the need for a threshold concentration of peptide in order to form an observable gel shift. This problem can be circumvented by measuring the association and dissociation rate constants for the bimolecular association reaction between RNA and protein, and calculating the equilibrium dissociation constant using the relationship

$$K_d = \frac{k_{off}}{k_{on}} \quad (5)$$

where k_{off} is the dissociation rate constant, and k_{on} is the association rate constant. Another important reason for measuring these rate constants is that in order for a minimized RNA substrate to serve as a good model for structural studies, it must display both

thermodynamic and kinetic stabilities similar to those of the wild type RNA (Long & Crothers, 1995).

Dissociation experiments were conducted by equilibrating a probe RNA with 50 nM BS15 for two hours at room temperature in the above conditions prior to adding 50-200 nM unlabeled Fr12 RNA (Figure 3.3), an RNA that is capable of binding BS15 with wild type affinity, to the reaction. Aliquots of 5 µl were taken at various time points, 1 µl of Type III Load buffer was added, and immediately loaded onto an electrophoresing gel. This method for determining the dissociation rate constant suffers from a potential artifact called facilitated dissociation. In some protein-nucleic acid interactions, the dissociation rate varies as a function of competitor nucleic acid that is first order with respect to the competitor concentration (Fried & Crothers, 1984; Weeks & Crothers, 1991). An alternative method which circumvents this problem is to dilute the binding reaction sufficiently such that the protein concentration falls well below the K_d. To measure the dissociation of a preformed complex by this method, the protein-RNA complex was formed as detailed above in a 20 µl reaction, and then dissociation was initiated by the addition of 980 µl of Binding buffer to the reaction. Aliquots of 10 µl were taken at various time points, 1 µl of Type III Load buffer was added, and immediately loaded onto an electrophoresing gel. The dissociation rate, k_{off}, was determined by fitting the data to

$$\frac{\Theta_t}{\Theta_{t=0}} = A\exp(-k_{off}t) \qquad (6)$$

using nonlinear least squares analysis.

Association rate experiments were conducted at room temperature under the above conditions by adding a specific amount of protein to radiolabeled probe RNA. At each timepoint, an 8 µl aliquot was taken, the reaction was quenched with the addition of 2 µl 5.5 mM unlabeled competitor RNA and 1 µl Type III Load buffer. The reaction was

allowed to quench for 30 seconds before loading onto a gel with the current on. The association kinetics were analyzed using the equation

$$\Theta_t = B + A\exp(-k_{on'}t) \qquad (7)$$

where $k_{on'}$ is the pseudo-first order rate constant of association. The pseudo-first order rate constant was measured at various protein concentrations, and the bimolecular rate constant, k_{on}, was calculated from the slope of the line through a graph of the pseudo-first order rate constant versus protein concentration.

RNA folding

The mobility of an nucleic acid species in a native polyacrylamide gel matrix is sensitive to its conformation as well as ligand binding, and thus has been used extensively to probe the conformation of DNA and RNA junction motifs (Duckett *et al.*, 1988; Clegg *et al.*, 1994; Bassi *et al.*, 1995). The primary variable in this experiment is the presence of divalent metal ions or organic polycationic species in the electrophoretic running buffer. The typical running buffer for this type of experiment is a solution of 0.045 M Tris-borate, pH 8.1 and 50 mM KCl, along with a variable polyvalent ionic species or EDTA. In order to prevent heating of the gel due to the presence of salt in the running buffer, these gels were run at 5-10 V/cm at 4 °C. To insure that the salt concentration in the gel did not change during electrophoresis, the buffer was recirculated between the upper and lower buffer chambers using a peristaltic pump at the rate of 50 ml/hr.

Chemical modification interference analysis

By exploiting the capability of an electrophoretic gel to separate unbound and bound RNA, another type of experiment can be performed to probe RNA moieties that are

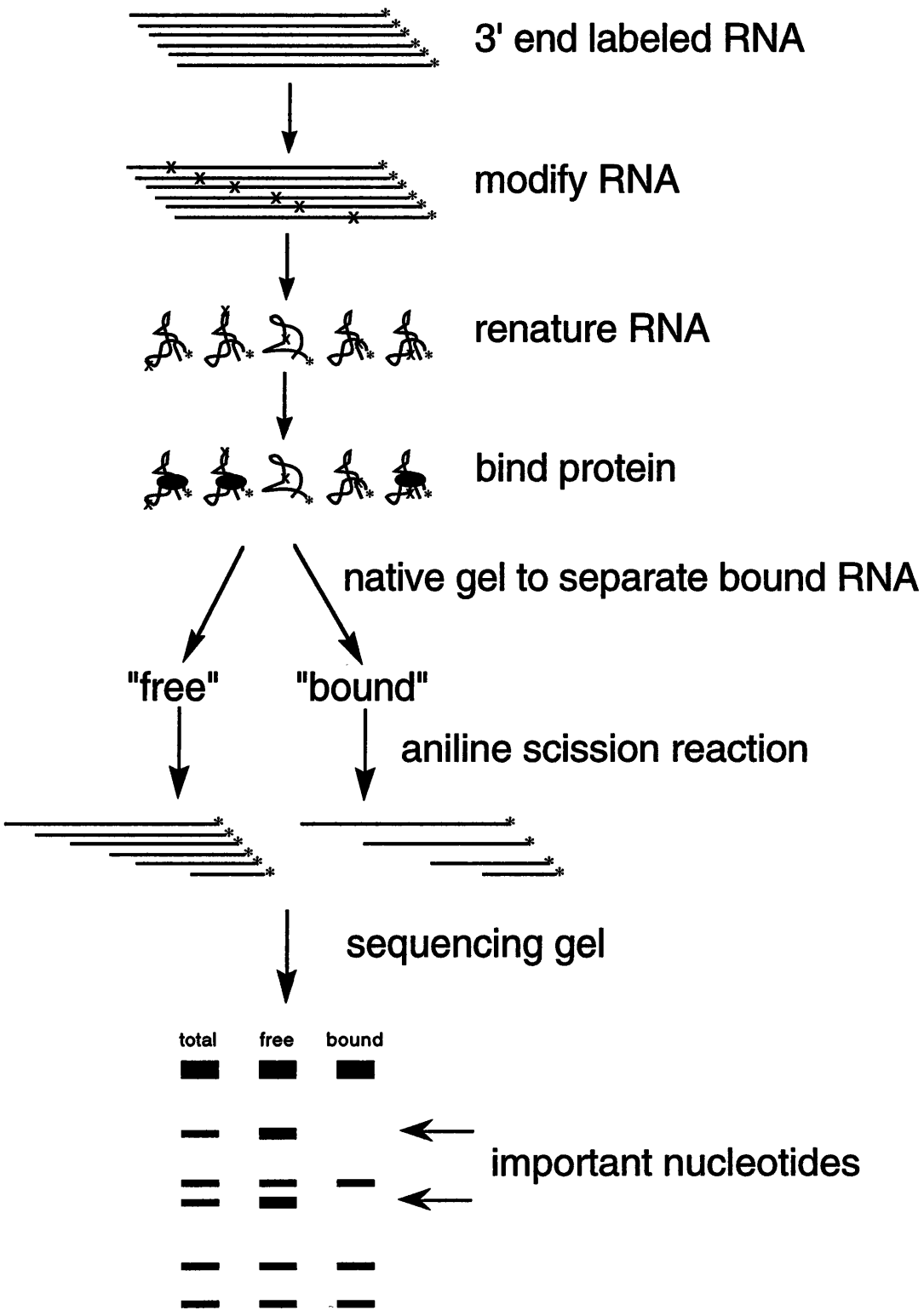


Figure 2.6: Experimental scheme for the modification interference assay.

important for function. Chemical modifications that disrupt the ability of a protein to specifically recognize a target nucleic acid can selected against by separating the modified RNA on the basis of its ability to bind protein using a native gel shift (experimental scheme illustrated in Figure 2.6). This approach yields in a single experiment results similar to those obtained using a complete set of point mutants (Conway & Wickens, 1987; Conway & Wickens, 1989). The Watson-Crick and Hoogsteen faces of the nucleotide bases can be probed to provide information about the chemical moieties that are necessary for formation of RNA structure as well as its interaction with other macromolecules. Probing with reagents that modify the phosphate backbone, such as ethylnitrosourea (ENU) (Calnan *et al.*, 1991), or the 2' hydroxyl group of the ribose sugar (Gaur & Krupp, 1993) generates additional information that cannot be accessed by alternative approaches such as mutagenesis.

The validity of this approach to analyze RNA-protein interactions is demonstrated by the comparison of the results of modification interferences experiments and structures obtained using crystallographic or NMR techniques. The interaction between human immunodeficiency virus (HIV) tat protein and TAR, an RNA hairpin containing a three nucleotide bulged loop located at the 5' end of the viral mRNA, has been demonstrated to be mediated by arginine residues in a basic region of the protein (Weeks *et al.*, 1990). Biochemical experiments have demonstrated that free argininamide, an arginine analog, serves as a reasonable model for the tat-TAR interaction (Calnan *et al.*, 1991; Long & Crothers, 1995). Chemical modification interference analysis of the interaction of peptides derived from the basic region of tat protein or argininamide with TAR RNA demonstrate that the critical components of the interaction are two base pairs above and below the trinucleotide bulge, a single uridine in the bulge, and several phosphates in the RNA backbone (Weeks *et al.*, 1990; Tao & Frankel, 1992). A model of the argininamide-TAR RNA interaction derived from NMR constraints positions the guanidinium group of argininamide contacting the phosphate backbone and a G-C base pair above the

trinucleotide bulge and a uracil in the bulge forming a triple with a A-U base pair above the bulge (Puglisi *et al.*, 1992). Modification interference analysis of the tat-TAR interaction supported a role for every functional group involved in the model of the argininamide-TAR interaction and its relationship to the tat-TAR interaction. More recent modification data generated using synthetic RNAs containing a single specific modification have largely supported this model of the mode of tat-TAR recognition (Hamy *et al.*, 1993).

A strong correlation between the information obtained from chemical modification interference analysis and NMR has also been demonstrated for the HIV rev protein-RRE RNA interaction. Biochemical analysis of a rev peptide-RRE RNA complex implicated an internal loop containing two unusual purine-purine base pairs as well as base pairs immediately flanking the bulge (Bartel *et al.*, 1991; Kjems *et al.*, 1992; Pritchard *et al.*, 1994). NMR studies of RRE RNA complexed with this peptide demonstrate that in the RNA all of the nucleotides and phosphates which are important for binding the peptide are brought into a conformation such that they surround a widened major groove centered around the internal loop (Battiste *et al.*, 1995). Additional NMR evidence indicates that the peptide lies in the widened major groove, flanked by the nucleotides and phosphates identified by chemical modification as important for complex formation (Battiste *et al.*, 1996). Similar results have been observed in the studies of U1A spliceosomal protein complexed with a small RNA hairpin derived from the U1 snRNA. Ethylation modification interference analysis of the protein-RNA interaction suggested that the U1A N-terminal RNP domain interacts with the 5' side of the hairpin loop and the 3' side of the hairpin stem (Jessen *et al.*, 1991). The recently solved crystal structure of the complex shows that two lysine residues which were determined to be important for the protein-RNA interaction by mutagenesis interact with the hairpin stem in the region implicated by ethylation modification interference (Oubridge *et al.*, 1994). However, it is also clear from the studies of the rev-RRE interaction and the U1A RNP-hairpin interaction that not every moiety identified need be directly involved in the recognition between the protein and RNA.

Some of the interferences observed in the rev peptide-RRE do not correspond to contacts, but rather interfere with ability of the RNA to form a stable structure (Battiste *et al.*, 1995). Thus, modification interference identifies residues on the RNA that are important either as a direct protein-RNA contacts or for formation of the proper RNA structure. Despite this limitation, detailed biochemical knowledge of an RNA-protein interaction can guide subsequent structural studies and provides a critical test for the validity of the resulting structural models.

The same reagents used for chemical sequencing of RNA were used as probes for chemical modification interference studies: diethyl pyrocarbonate (DEPC) to probe the purine N7 positions, and hydrazine to probe pyrimidine bases. In addition, the phosphate backbone of the RNA was probed using ethylnitrosourea (ENU), an N-nitroso alkylating agent that ethylates the oxygen atoms of the phosphodiester linkages, as shown in Figure 2.7. The resulting phosphotriester is highly unstable and readily hydrolyzes under moderately basic conditions. The one caveat to using this reagent is that the phosphotriester linkages are so unstable that they slowly cleave under the conditions used to separate the free and bound RNA in a polyacrylamide gel. Thus, a significant amount of degradation of the RNA is observed during this step, particularly in the free RNA, which can make analysis of the modification interference data difficult. Also, it is important to perform ENU modification interference analysis using both 5' and 3' end-labeled RNA, since the ENU modification ladder is prone to band compression artifacts around sites of stable RNA secondary structures such as tetraloops. Unlike aniline strand scission, phosphotriester cleavage by hydroxide leaves a 5' hydroxyl group and a 3' phosphate, allowing for both of the resultant fragments to be analyzed on a sequencing gel.

Chemical modification interference was performed with approximately 50 to 250 ng ($2x10^6$ cpm) of 3' end-labeled RNA, and modified with approximately one hit per molecule using diethyl pyrocarbonate (DEPC) (ICN Biochemical) or hydrazine (Sigma) as described above. The same quantity of 5' or 3' end-labeled RNA was modified with

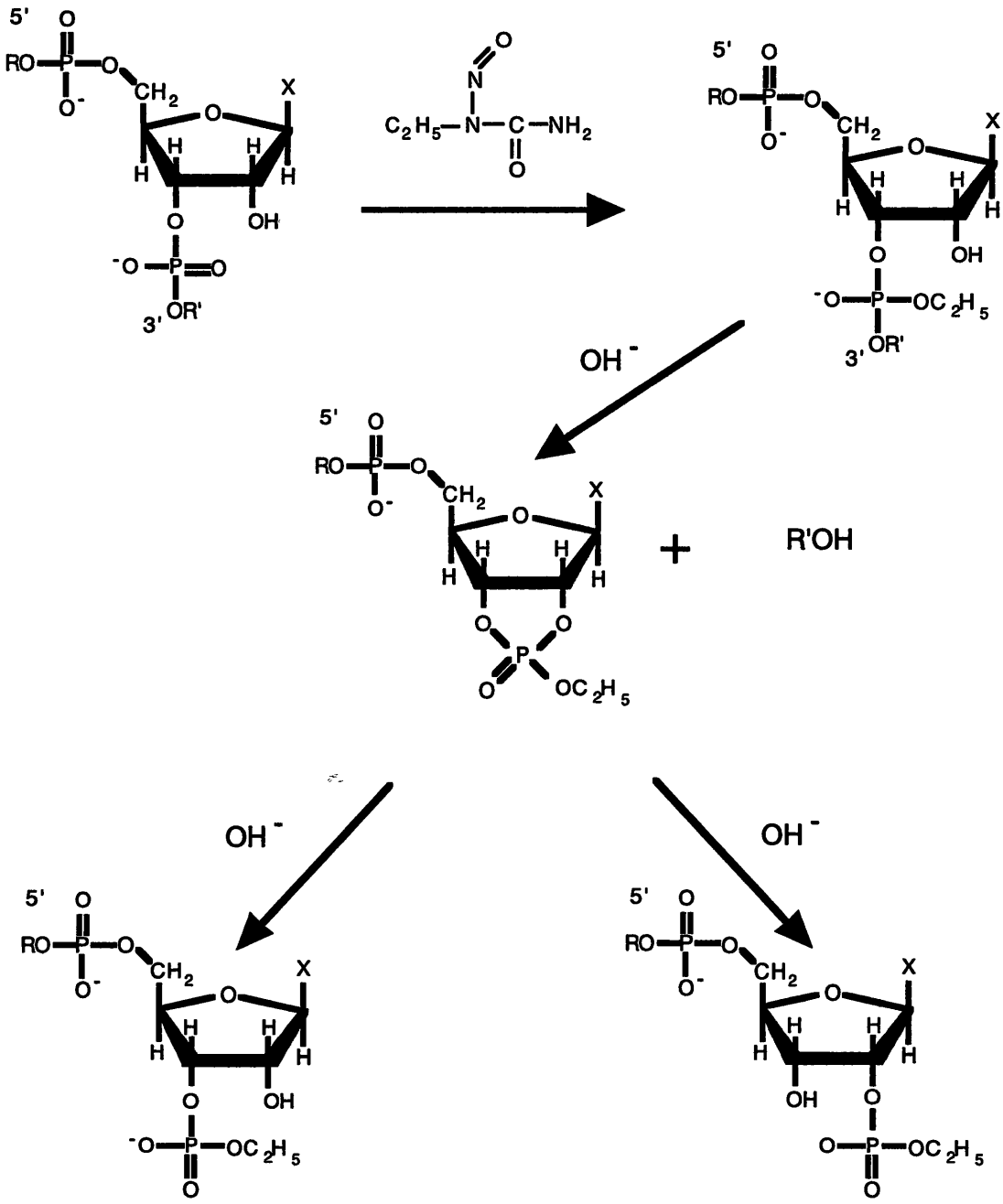


Figure 2.7: Chemical modification of RNA by ethylnitrosourea. This chemical reacts with the phosphate backbone to yield a phosphotriester linkage at the modified position, which is susceptable to cleavage upon alkaline treatment.

ethylnitrosourea (ENU) (Sigma) according to Calnan *et al.* (Calnan *et al.*, 1991). The RNA was dissolved in 10 µl of ENU buffer (150 mM sodium cacodylate, pH 8.0, 1 mM EDTA), 2 µl 1 mg/ml unfractionated yeast tRNA, and sufficient deionized H_2O to yield a total volume of 20 µl, and chilled on ice. To this reaction, 5 µl of freshly made saturated ENU in 100% ethanol was added, vortexed, and incubated for 30 seconds at 90 °C. This reaction was quenched by the addition of 20 µl 0.3 M sodium acetate, pH 5.2, and 200 µl 100% ethanol, the RNA precipitated by incubating in a dry ice/ethanol bath for 10 minutes, and pelleted. The RNA was resuspended in 200 µl 0.3 M sodium acetate, pH 5.2, precipitated by adding 800 µl 100% ethanol, and pelleted. After the pellet was washed with 70% ethanol and dried for 5 to 10 minutes in a Speed-Vac, the ENU modified RNA was ready for modification interference.

The modified RNA was resuspended in Binding buffer (10 mM K-HEPES, pH 7.5, 50 mM potassium acetate, 0.1 mM EDTA) and renatured by heating to 90 °C for one minute and chilling on ice for 3 minutes. The rest of the binding reaction components were added to the modified RNA along with a sufficient concentration of BS15 to bind 40-60% of the RNA. The samples were incubated for one hour at 25 °C and one hour at 4 °C prior to loading onto a native polyacrylamide gel. Fr1 RNA was run on an 8%, 29:1 acrylamide:bisacrylamide gel at 25 V/cm at 4 °C for approximately four hours; Fr7 and Fr15 RNA were run on a 20% 19:1 acrylamide:bisacrylamide gel at 25 V/cm at 4 °C for 16 hours. The free and bound RNA was visualized by exposing the gel to autoradiographic film for one hour, and then excised and electroeluted from the gel using an Elutrap (Schleicher & Schuell). The eluted RNA was precipitated with ethanol twice and washed with 70% ethanol prior to cleavage of the modified sites with aniline for DEPC or hydrazine modified RNA as described previously. ENU modified RNA was cleaved by dissolving the RNA pellet in 20 µl of 0.1 M Tris-Cl, pH 9.0 and incubating for 15 minutes at 50 °C. (Calnan *et al.*, 1991). The samples were analyzed on an 8% polyacrylamide/8 M

urea denaturing gel. ENU reactions were run next to RNase T_1 and alkaline hydrolysis sequencing ladders in order to identify the bands in the ENU lanes.

Modifications of the Watson-Crick base pairing face of the RNA bases are not capable of being detected by strand scission reactions; in order to detect these modifications, the RNA needs to be reverse transcribed. During cDNA synthesis by reverse transcriptase, any modification in the RNA that prevents a deoxyribonucleotide triphosphate from base pairing to the RNA will not be incorporated, and thus, modifications of the Watson-Crick face of the RNA terminate reverse transcription. The cDNA synthesized from the pools of free and bound RNA can be electrophoresed on a sequencing gel and the sites of modification interference determined.

Several chemicals are capable of specifically modifying the Watson-Crick face of RNA bases. Dimethyl sulfate, along with methylation of the N7 position of guanines, modifies the N1 position of adenine and the N3 position of cytidine to a much lesser extent (Figure 2.8). Since methylation of the N7 position of guanine does not terminate reverse transcription, these modifications are not visible by this method. 1-cyclohexyl-3-(2-morpholinoethyl)carbodiimide metho-*p*-toluenesulfonate (CMCT) modifies the N3 position of uracil and the N_1 position of guanine to a slightly lesser extent (Ehresmann *et al.*, 1987), as illustrated in Figure 2.9.

In order to produce an RNA fragment that could be utilized in a reverse transcription-based assay, pB16SFr15 (see Appendix 1) was digested with *Eae*I (New England Biolabs), which cleaves seventeen nucleotides downstream of the *Sma*I site at the 3' end of the Fr15 RNA gene. Fr15/*Eae*I RNA was transcribed, purified, and 3' end-labeled in the usual fashion; this RNA contained a single-stranded 3' tail, $^{5'}$GGGAAGCUUGGCACUGGCC, that served as a site for hybridization of a complementary oligonucleotide for reverse transcription. Prior to using this RNA for modification interference assays, the ability of the RNA to bind BS15 was tested by an equilibrium binding assay as monitored using a gel mobility shift assay, and Fr15/*Eae*I

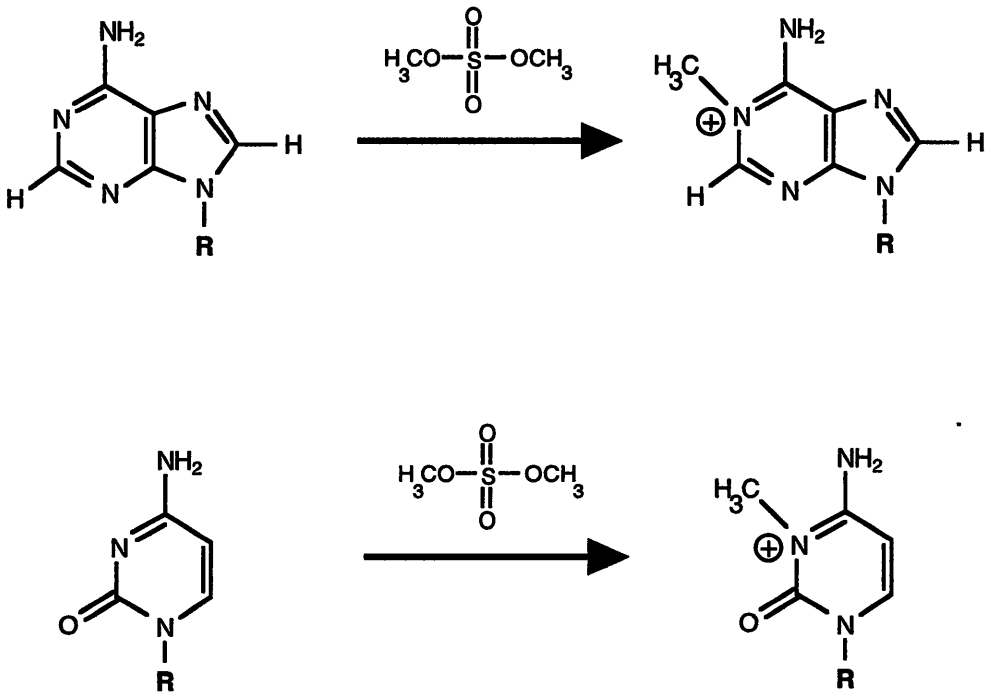


Figure 2.8: Modification of the Watson-Crick face of the adenine and cytidine bases of RNA with DMS.

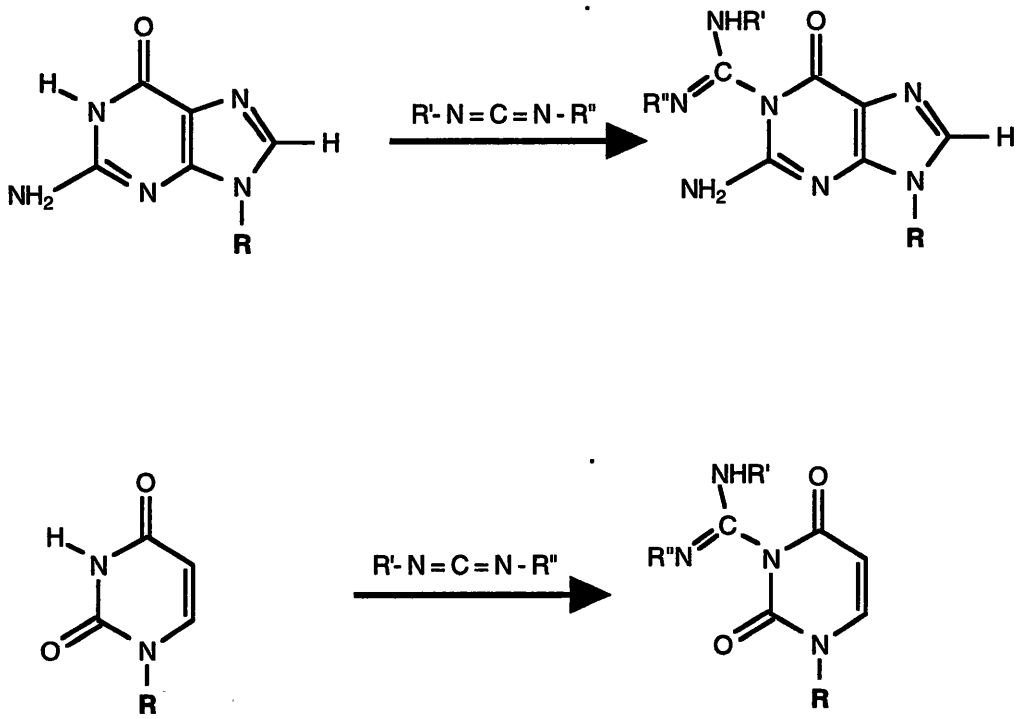


Figure 2.9: Modification of the Watson-Crick face of the guanine and uracil bases with CMCT.

RNA was found to bind with wild type affinity. Therefore, the 3' single stranded tail does not interfere with the RNA's ability to form the proper secondary structure or its ability to interact with BS15. Approximately 1.25 μg Fr15/*Eae*I RNA (supplemented with 10^6 cpm of 3' end-labeled RNA) was modified with approximately one hit per molecule using a final concentration of 1.9 mg/ml CMCT (Aldrich) under denaturing conditions at 90 °C in 50 mM sodium borate, pH 8.0, 1 mM EDTA for one minute (Mougel *et al.*, 1987). The reaction was quenched by ethanol precipitation and the resulting RNA pellet was washed with 70% ethanol. Modification of 1.25 μg Fr15/*Eae*I RNA (supplemented with 10^6 cpm of 3' end-labeled RNA) with DMS (Sigma) was accomplished as described (Peattie, 1979). The modified RNA was incubated with BS15 such that approximately 50% of the RNA was complexed under the conditions described above. Free and bound RNA were separated on a 20%, 19:1 acrylamide:bisacrylamide gel at 25 V/cm at 4 °C for 16 hours, and extracted from the gel by electroelution. The modified RNA was probed by reverse transcription using a 5' labeled DNA oligonucleotide complimentary to the 3' tail of Fr15/*Eae*I RNA using a modification of the primer extension method of Moazed *et al.* (Moazed *et al.*, 1986; Recht *et al.*, 1996). 2.5 μg of Fr15/*Eae*I RNA and 0.04 μg of the DNA primer was hybridized in 1x Hybridization buffer (50 mM K-HEPES, pH 7.0, 100 mM KCl) for two minutes at 85 °C and ramped down to 45 °C over ten minutes in a thermocycler. Extension was performed by the addition of 0.38 mM of each dNTP, and 2 units of avian myeloblastosis virus (AMV) reverse transcriptase (Seikagaku) in Extension buffer (130 mM Tris-HCl, pH 8.5, 10 mM $MgCl_2$, 10 mM DTT). The reactions were incubated at 42 °C for one hour prior to quenching. Sequencing reactions using unmodified RNA contained 0.64 μM of a dideoxynucleotide and 3.1 μM of each dNTP, and 2 units AMV reverse transcriptase in 1x Extension buffer. These reactions were incubated for 30 minutes at 42 °C and then supplemented to 1.3 μM of a dideoxynucleotide and 83 μM of each dNTP and incubated for another 30 minutes. All extension reactions were quenched by ethanol precipitation for 10 minutes at room temperature. The pellets

were resuspended in 20 µl 50 mM Tris-HCl, pH 8.0, 0.5% sodium dodecyl sulfate, and 7.5 mM EDTA. The RNA was destroyed by the addition of 3.5 µl 3 M KOH and incubation at 90 °C for three minutes and at 37 °C for 16 hours. The reaction was neutralized by adding 6 µl glacial acetic acid, and ethanol precipitated. The cDNA was washed with 70% ethanol, dried, and resuspended in 10 µl of Formamide stop mix (95% formamide, 20 mM EDTA, 0.025% bromophenol blue, 0.025% xylene cyanol) and loaded onto a denaturing 12%, 29:1 acrylamide:bisacrylamide gel. The polyacrylamide gel was run for two hours and visualized by autoradiography.

Chemical Footprinting

Protein-RNA interactions can be directly probed in solution by a number of footprinting techniques. The most common technique is nuclease footprinting using a combination of the RNases described previously. Although these can be readily employed under native conditions, they suffer from the problem of being bulky probes, and thus subject to significant steric hindrance from the protein. Thus, the observed footprint for a protein bound to an RNA is often much more extensive than the site of direct contact. Also, most nucleases are probes for unstructured single stranded regions, and if the protein binds predominantly highly structured, double stranded regions of RNA, it will be difficult to visualize a substantial footprint. Using chemical footprinting overcomes this first problem, but many chemicals still preferentially attack bases in unstructured regions of the RNA. Diethyl pyrocarbonate, for example, is completely unreactive to purine N7 positions in the major groove of double-stranded *A*-form RNA (Peattie & Gilbert, 1980; Van Stolk & Noller, 1984; Weeks & Crothers, 1991). Furthermore, DEPC modifies the imidazole ring of histidine (Creighton, 1984), and thus, can inactivate RNA-binding proteins in which this functional group is important. Other chemical footprinting techniques using ethylnitrosourea (ENU) or Fe(II)-EDTA do not suffer from these problems. These

chemicals modify RNA moieties in both single and double stranded regions; only those functional groups that are solvent inaccessible due to protein binding or RNA tertiary structure are protected from attack. However, at low temperatures, the probes react very slowly, resulting in long incubation periods in order to generate sufficient levels of modification.

The footprinting technique employed for these studies utilizes the ability of molecular iodine to cleave RNA at sites where a phosphorothioate has been substituted in the RNA backbone (Figure 2.10) (Gish & Eckstein, 1988; Schatz *et al.*, 1991). In this technique, the RNA is synthesized in the presence of a sufficient concentration of ribonucleoside 5'-O-(1-thiotriphosphates) to generate approximately one incorporation per RNA molecule. Since the modifications are incorporated during synthesis, separate RNAs can be in which only one of the four phosphorothioates is incorporated, allowing for unambiguous sequencing of the footprinting reaction. Low concentrations of molecular iodine (100 μM to 1 mM) rapidly cleave the RNA backbone specifically at the phosphorothioate substitution, so reactions times are typically very short (about one minute). This reagent can also be easily quenched by the addition of thiourea to the reaction. The only limitation of this experiment is that if the substitution of a sulfur atom at a particular position abolishes the binding of the protein to the RNA, then this position will not appear to be protected by the protein. However, phosphorothioates that interfere with protein binding can be revealed using a modification interference approach.

The synthesis of phosphorothioate RNA was performed under the standard *in vitro* RNA transcription conditions using plasmid templates described in the Appendix 1, except that the NTPs were supplemented with 0.31 mM ATP[αS](Amersham) and UTP[αS], or 0.31 mM CTP[αS] and GTP[αS] (Schatz *et al.*, 1991). These RNAs were purified and labeled at the 3' terminus with ^{32}P using [5'-^{32}P]pCp (England & Uhlenbeck, 1978). Prior to each experiment, the RNAs were heated to 90 °C for one minute and cooled on ice for three minutes to melt any potential aggregates or misfolded RNA. Reactions of RNA

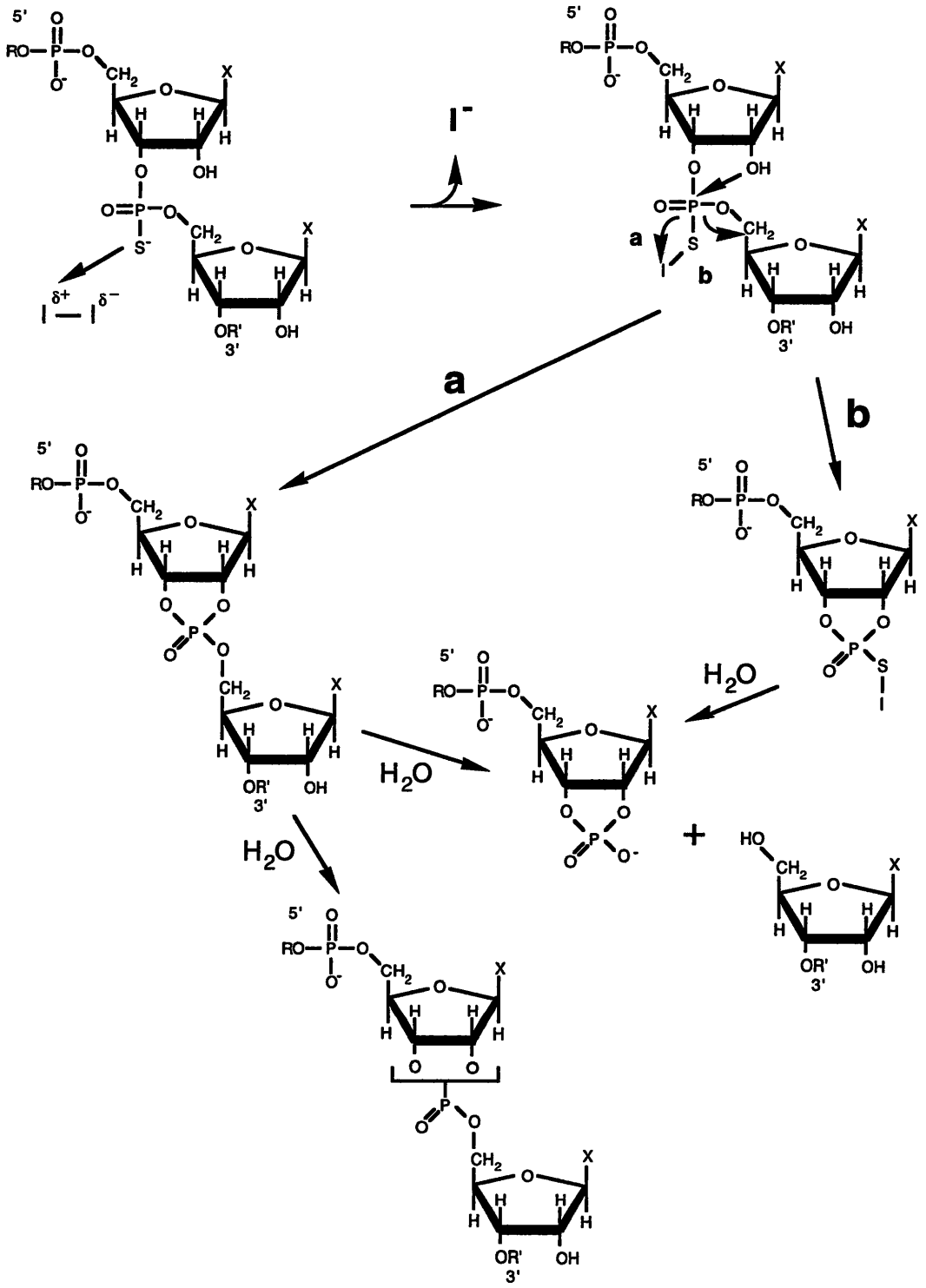


Figure 2.10: Cleavage reactions of phosphorothioate labeled RNA with molecular iodine.

alone or RNA with 200 nM BS15 were incubated in 20 µl reaction volumes under the conditions used for native gel electrophoresis experiments. After incubation of the RNA for 1 hour at 25 °C and 1 hour at 4 °C, the RNA was probed by the addition of 2 µl of an iodine solution to give final iodine concentrations of 100 µM, 500 µM, or 1 mM and incubated on ice for one minute (Schatz *et al.*, 1991). A denatured RNA control was incubated at the same time for one minute at room temperature immediately after being heated to 90 °C for three minutes. The reactions were quenched by the addition of sodium acetate, pH 5.6, to a final concentration of 0.3 M, extracted with 50 µl of phenol, and precipitated with 100% ethanol. The precipitated RNA was washed with 70% ethanol, dried, resuspended in 10 µl of Urea stop mix, and loaded onto a sequencing 8%, 29:1 acrylamide:bisacrylamide denaturing gel. The polyacrylamide gel was run for two hours at 60 watts, constant wattage, and visualized by autoradiography.

The dimerization question

Structural studies of RNA oligonucleotides using NMR spectroscopy require samples in which the RNA is present in millimolar concentrations. At these high concentrations, many RNAs exhibit a strong tendency to form dimers or other higher order aggregates. Hairpin structures, by virtue of their high degree of self-complimentarity, display an equilibrium between the monomeric stem-loop structure and a dimeric duplex containing an internal loop. Thus, prior to studies that require high RNA concentrations, the behavior of the RNA must be explored using techniques that can address the molecularity of the RNA structure. Several techniques are capable of addressing this issue. The melting of RNA structure can be monitored by observing the temperature dependent UV absorbance. For a monomeric RNA species, the midpoint of the thermal melting transition, T_m, will be independent of concentration, while dimeric RNA species exhibit a concentration dependence to their thermal stability (Puglisi & Tinoco, 1989). Thus, by

observing the UV absorbance melting curve of an RNA over a range of concentrations, it can be determined whether the RNA displays a tendency to multimerize at high concentrations. Native gel electrophoresis can also resolve RNAs in different multimeric species (Henderson *et al.*, 1987); using conditions similar to those described in previous sections, a range of RNA concentrations can be subjected to gel electrophoresis to determine whether it has a tendency to multimerize.

The favored technique for the determination of the molecularity of RNA species under a variety of conditions is size exclusion chromatography. This form of chromatography separates molecules on the basis of their apparent molecular volume using a porous matrix. Large molecules do not have access to the inside of the porous matrix, and thus are eluted from the column before small molecules that are capable of entering the matrix, and are thus retained. If the exchange between the monomeric and dimeric species is slow compared to the amount of time they spend in the column, they can be separated (Puglisi, 1989). However, because of the exchange between the monomeric and dimeric forms of the RNA during chromatographic separation, particularly due to diffusion of the RNA bands in the column, which favors monomer formation, size exclusion chromatography should only be used as a qualitative tool for the analysis of the monomer-dimer equilibrium.

Size exclusion chromatography was performed on a high performance liquid chromatographic system (Beckman) using a 300 x 7.8 mm Bio-Sil SEC 125-5 (Bio-Rad) analytical column for RNAs of less than 60 nucleotides, and a 300 x 7.8 mm Bio-Sil SEC 400-5 (Bio-Rad) analytical column for RNAs of greater than 60 nucleotides. Prior to application of an RNA, the column was equilibrated for 15 minutes with an elution buffer at a flow rate of 1 ml/min at room temperature. The solvent typically contained 10 mM sodium phosphate, pH 6.5, 100 mM sodium chloride, and either 10 mM magnesium chloride or 1 mM EDTA. RNA was injected onto the column at various concentrations ranging from 10 μM to 1 mM in a volume of 2 to 5 μl, followed by a chase of 10 μl of

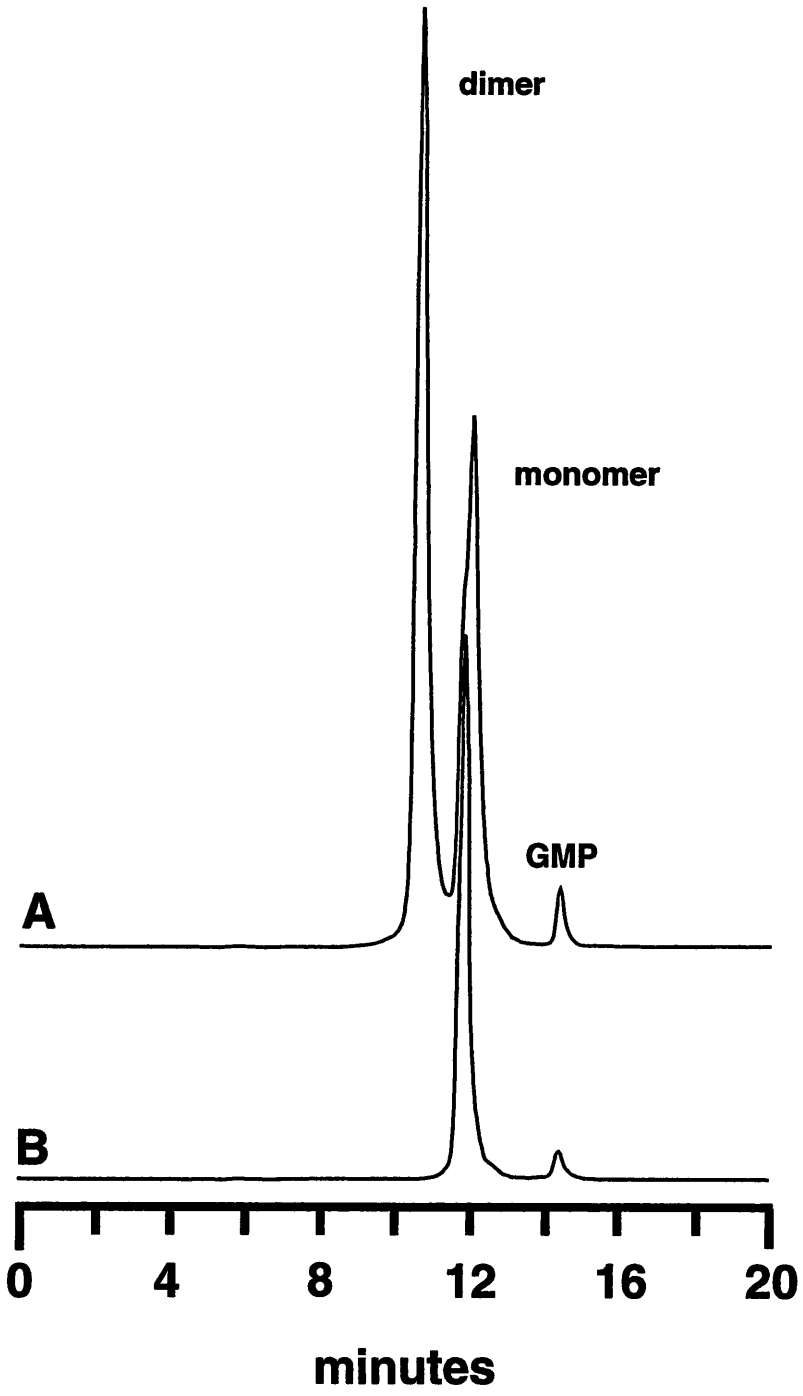


Figure 2.11: Size exclusion HPLC of RNA fragments. Trace A shows an RNA that at 1 mM concentration is predominantly dimerized, whereas trace B shows an RNA that is completely momomer at the same concentration. The small GMP peak is to indicate the time at which the smallest molecules elute from this column (elution condtions are described in the text).

elution buffer. The RNA species were detected by monitoring their UV absorbance at 260 nm. A typical elution profile of an RNA species that dimerizes at higher concentration is shown in Figure 2.11.

Nuclear Magnetic Resonance (NMR) Spectroscopy

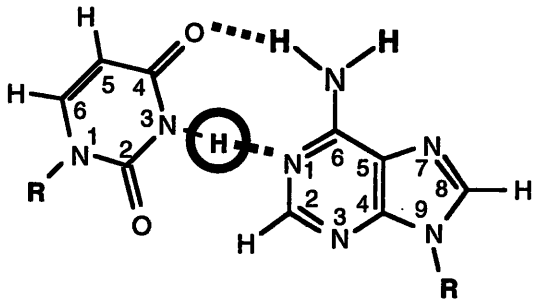
There are two principal tools for the determination of macromolecular structure with atomic-level detail: X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. Each of these methods has had a significant impact upon our knowledge of RNA structure in recent years. Using X-ray crystallography, the structures of large functional RNAs such as tRNA (Jack *et al.*, 1976; Schevitz *et al.*, 1979; Moras *et al.*, 1980; Woo *et al.*, 1980), the hammerhead ribozyme (Pley *et al.*, 1994b; Scott *et al.*, 1995b), and the P4/P6 domain of the group I intron have been determined (Cate *et al.*, 1996), as well as a number of protein-RNA complexes (Rould *et al.*, 1989; Ruff *et al.*, 1991; Biou *et al.*, 1994; Oubridge *et al.*, 1994; Valegard *et al.*, 1994; Nissen *et al.*, 1995). Along with these structures, new techniques have been developed for synthesizing and purifying RNA suitable for crystallography (Price *et al.*, 1995; Ferré-D'Amaré & Doudna, 1996) and screening procedures for obtaining diffraction quality RNA crystals (Doudna *et al.*, 1993; Scott *et al.*, 1995a; Anderson *et al.*, 1996). An alternative technique that has proved to be equally robust for determination of small RNA structures is NMR spectroscopy, which has used to solve the structure of stable RNA loop structures (Cheong *et al.*, 1990; Heus & Pardi, 1991; Jucker & Pardi, 1995; Dieckmann *et al.*, 1996; Fan *et al.*, 1996; Yang *et al.*, 1996), RNA aptamers (Dieckmann *et al.*, 1996; Fan *et al.*, 1996; Yang *et al.*, 1996), and RNA-peptide complexes (Puglisi *et al.*, 1995; Battiste *et al.*, 1996). This in large part has been facilitated by the development of techniques to enzymatically synthesize large quantities of RNA (Milligan *et al.*, 1987), label RNA with NMR-active nuclei such as ^{13}C and ^{15}N (Batey *et al.*, 1992; Nikonowicz *et al.*, 1992; Hines *et al.*,

1993), and new experimental techniques to address specific aspects of RNA structure (Farmer *et al.*, 1993; Sklenar *et al.*, 1993; Farmer *et al.*, 1994; Heus *et al.*, 1994; Marino *et al.*, 1994; Sklenar *et al.*, 1994). A detailed review of the entire process by which the solution structure of an RNA fragment is determined is given by Varani *et al.* (Varani *et al.*, 1996), and for the sake of brevity, only those issues relevant to the development of an S15-rRNA complex amenable to NMR studies as detailed in this thesis will be discussed below.

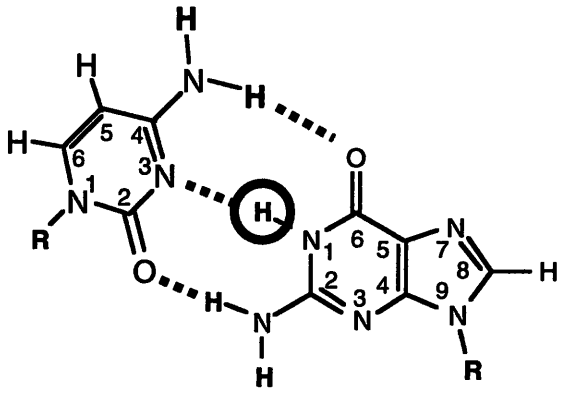
NMR applied to RNA

The fundamental observable of NMR spectroscopy is the resonances of all of the protons in the molecule of interest (for a general review of the relevant physics and practical aspects behind the NMR phenomena, the reader is referred to Wüthrich, 1986 and Derome, 1987). In RNA (as well as most other types of biological macromolecules) there are two types of observable protons: *exchangeable* protons, which are labile and can be exchanged with solvent protons, and *non-exchangeable* protons that cannot undergo exchange with solvent. Protons that fall into these two classes in RNA are illustrated in Figure 2.12.

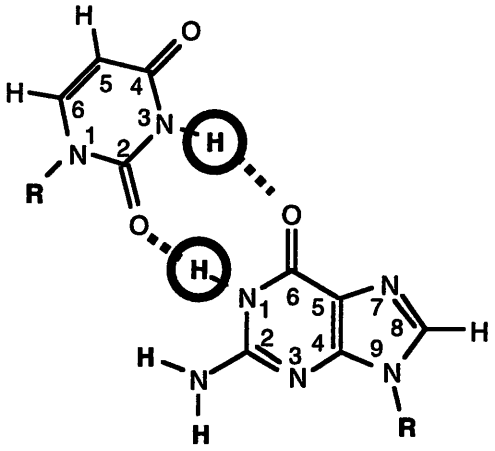
The non-exchangeable spectrum of a typical RNA, as observed in 99.996% D_2O such that all of the labile protons have been exchanged for deuterons, is shown in Figure 2.13. These protons fall into three distinct regions based upon their observed chemical shift. Chemical shift (δ) defines the location of a particular proton resonance along the frequency axis (which is generally converted into parts per million (ppm) units) and is exquisitely sensitive to the chemical environment of the proton in the macromolecule. The aromatic protons found in the nucleotide bases, which typically resonate between 7.0-8.5 ppm (except for the pyrimidine H5 resonance, which is found between 5.0-6.0 ppm) and the ribose anomeric protons (H1'), found between 5.0-6.5 ppm, are the most useful for determination of structure using two-dimensional (2D) homonuclear NMR. In contrast, the ribose H2', H3', H4', H5'/5" protons all resonate between 4.0-5.0 ppm, and the extreme



U-A Watson-Crick pair



C-G Watson Crick pair



U-G wobble pair

Figure 2.12: Commonly found base pairs in RNA. Protons that are readily exchangeable with bulk solvent are in boldface, and imino protons, which are NMR-observable only when they are involved in a base pair, are circled.

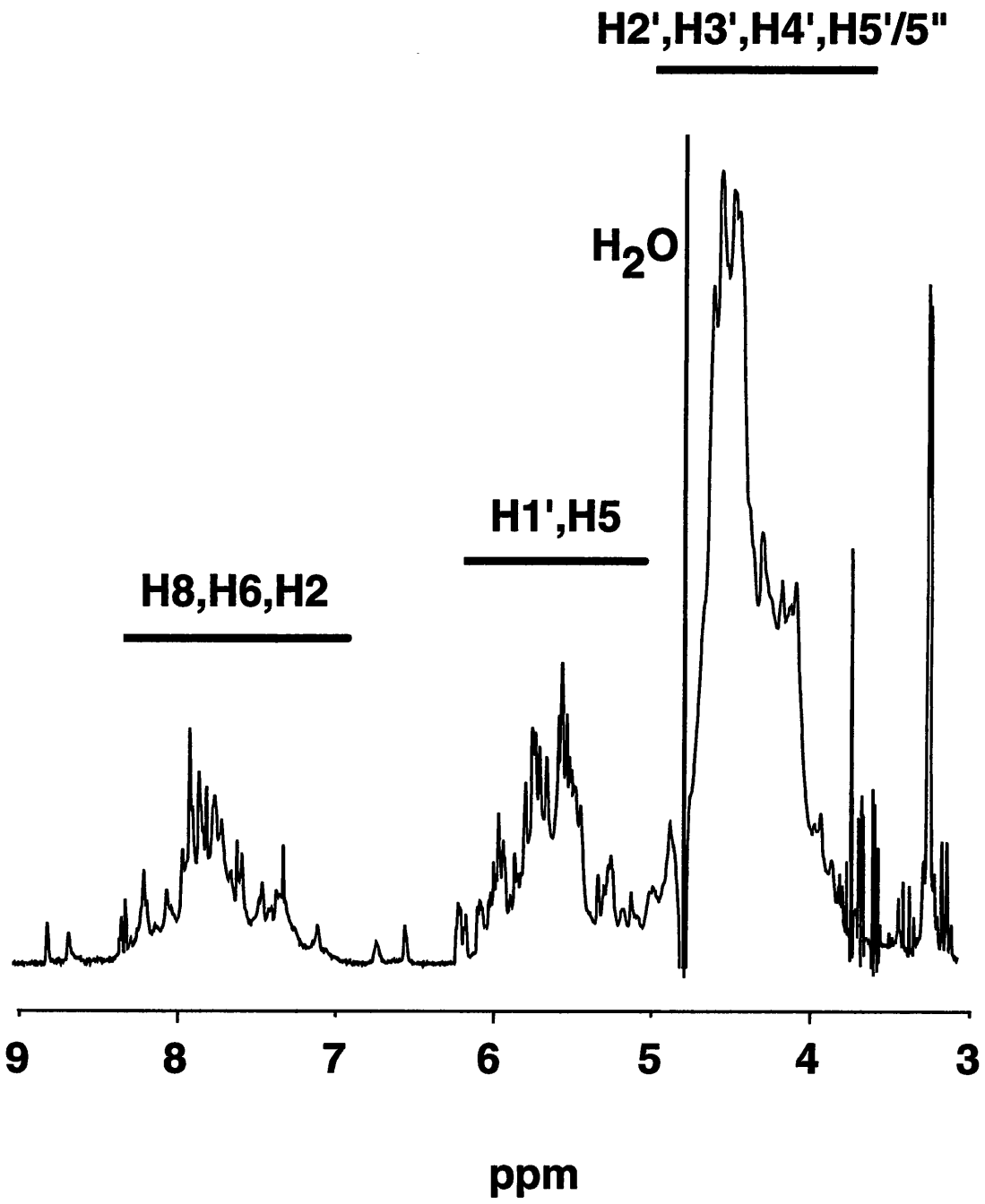


Figure 2.13: One-dimensional spectrum of RNA in D_2O, displaying the non-exchangeable protons. Regions in which protons in RNA typically resonate are indicated above the spectrum by solid bars.

degree of crowding observed in this region of the spectrum in 2D-NMR experiment makes them more difficult to assign. Many recent advances in RNA NMR have been specifically designed to overcome this fundamental problem. The exchangeable spectrum of RNA, as shown in Figure 2.14, is observable in an aqueous solution containing 90% H_2O/10% D_2O (a minimal amount of D_2O is required in the sample for technical reasons). The imino protons of guanosine and uracil, which resonate far downfield of all other protons (between 10-14.5 ppm) and are typically well-resolved in medium-sized RNA fragments (around 25-35 nucleotides) studied by NMR, and are diagnostic of secondary structure. Only protons that are involved in base pairs will be protected from rapid exchange with bulk solvent, and thus observable. Since there is one imino proton per Watson-Crick pair (or two for a wobble G-U pair), a simple one-dimensional (1D) spectrum can be indicative of whether the RNA is properly structured under the conditions (pH, salt concentration, temperature, and macromolecule concentration) used for study.

The basic experiments utilized for the observation and assignment of proton resonances in RNA take advantage of two fundamental phenomena: the nuclear Overhauser effect and scalar coupling. The nuclear Overhauser effect (nOe) is due to the direct magnetic interaction (dipolar coupling) through space between two different nuclei. The intensity of this effect between two isolated spins is described by the general equation

$$\text{NOE} \propto \left\langle \frac{1}{r^6} \right\rangle \bullet f(\tau_c) \qquad (8)$$

where r is the distance between the two nuclei and τ_c is the correlation time of the macromolecule Because of the $1/r^6$ dependence on the interproton distance, this effect can be used to determine through-space distances between protons; in practical application, an nOe can only be observed between two protons that are less than 5 Å apart. The dipolar

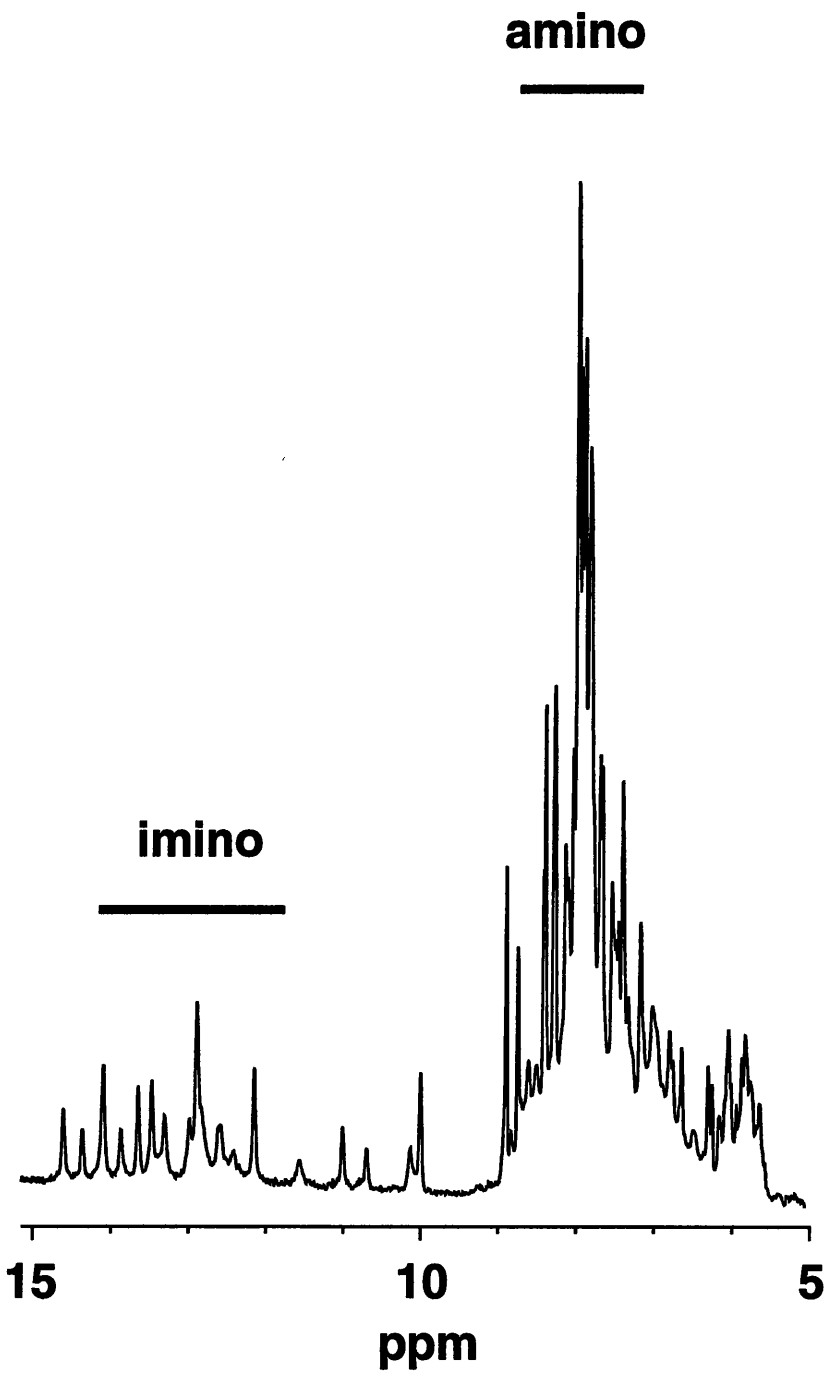


Figure 2.14: One-dimensional spectrum of RNA in H_2O/D_2O, highlighting the exchangeable imino and amino protons in RNA. Solid bars indicate the regions in which the exchangeable protons typically resonate.

coupling between spins is observed using a 2D experiment called nuclear Overhauser effect spectroscopy (NOESY), which is typically used to make sequence specific assignments of the proton resonances and to determine interproton distances in the macromolecule of interest.

Another 2D experiment, J-correlated spectroscopy (COSY), takes advantage of the through-bond correlation between two protons which are connected by three chemical bonds or less. The magnitude of the coupling constant (J) between two protons is determined in part by the torsion angle between two vicinal protons as described by a Karplus relation, and thus can be used to determine molecular conformation. In RNA, the ribose sugar adopts one of two primary conformations: C2'-*endo* and C3'-*endo* (Figure 2.15). In the C3'-*endo* conformation, typically occupied by ribose sugars that are in *A*-form helical segments of RNA, have a small $J_{H1'-H2'}$ (1-3 Hz) and are thus the correlation between these two protons in unobservable in a COSY experiment. Residues in non-*A*-form regions of an RNA, however, often adopt C3'-*endo* conformations with a large $J_{H1'-H2'}$ coupling constant (8-10 Hz), and are readily observable. Protons which are connected by four chemical bonds or more are capable of being correlated in a variation of the COSY experiment, total correlation spectroscopy (TOCSY). Through bond experiments are particularly useful in making sequence-specific assignments in macromolecule since assignment strategies based upon through-bond experiments are independent of model-based assumptions, unlike assignment procedures using NOESY experiments.

Two problems that are often encountered in NMR spectroscopy of RNA, spectral crowding and model-biased assignment procedures, can be overcome by using heteronuclear NMR techniques. The fundamental 2D heteronuclear experiments, heteronuclear single- and multiple- quantum coherence spectroscopy (HSQC and HMQC) correlate the spin of a proton to that of an NMR-active heteroatom (^{13}C or ^{15}N) to which it is directly attached. This is particularly useful for the assignment of proton resonances in

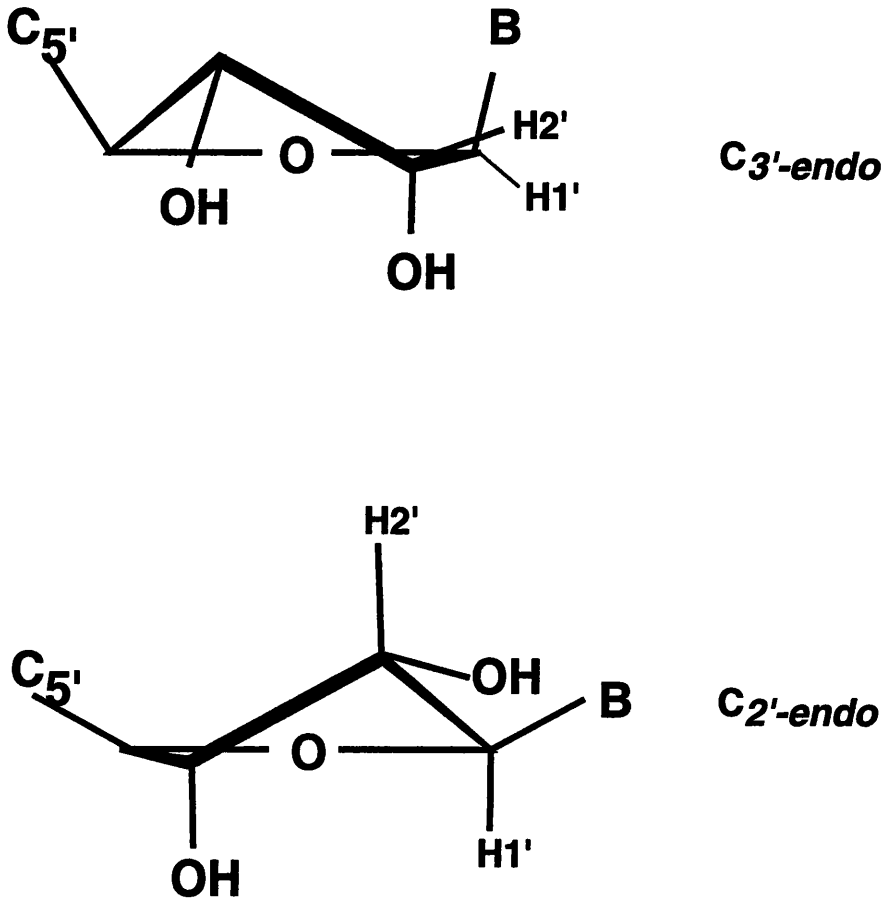


Figure 2.15: The two predominant sugar puckering modes found in RNA. The $C_{3'\text{-}endo}$ conformation, which dominates in a standard A-form helix, has a dihedral angle between protons H1' and H2' that is close to 90°, which yields a small coupling constant between these two protons ($J_{H1'\text{-}H2'}$ ~ 1-2 Hz). The $C_{2'\text{-}endo}$ conformation is generally found in non A-form regions of RNA and has a dihedral angle between protons H1' and H2' that is close to 180°, yielding a large coupling constant between these two protons ($J_{H1'\text{-}H2'}$ ~ 7-8 Hz).

the ribose sugar, which display a high degree of spectral overlap in homonuclear experiments, but whose carbon resonances are well resolved. This experiment is more commonly used in combination with the 2D NOESY, COSY, and TOCSY homonuclear experiments to create three- and four-dimensional experiments such as NOESY-HSQC and HCCH-TOCSY, which is critical for the interpretation of the spectra of the larger macromolecules studied by NMR (20-40 kDa). These and other experiments designed specifically for observing RNA have led to the development of techniques for assigning proton resonances using through-bond experiments only, which does not rely upon assumptions about the structure of the RNA.

Chemical exchange and relaxation processes in NMR

The quality of spectra that can be obtained for a macromolecule is dependent upon the observed resonance linewidths, which is turn, dependent upon many factors. One process that can result in broadened linewidths that is often observed in the NMR spectra of RNA is exchange. There are two types of exchange processes in NMR: chemical exchange (the physical exchange of protons with the bulk solvent, such as imino and amino protons) and conformational exchange (such as a ribose sugar flipping between the C2'-*endo* and C3'-*endo*) conformations (Wüthrich, 1986). As an example, consider two equally occupied states, A and B, with a proton that exchanges locations as characterized by the frequencies (in Hertz) ν_A and ν_B. If the rate of exchange, k_{AB}, between the two states is rapid such that $k_{AB} >> 2\pi(\Delta\nu_{AB})$, a situation called *fast exchange*, a single narrow resonance will be observed, whose observed frequency is the average of the frequencies of the proton in each state, as illustrated in Figure 2.16. The other limiting situation, *slow exchange*, occurs when $k_{AB} << 2\pi(\Delta\nu_{AB})$, and gives rise to spectra in which two narrow resonances are observed for the proton, one for state A and one for state B. In between

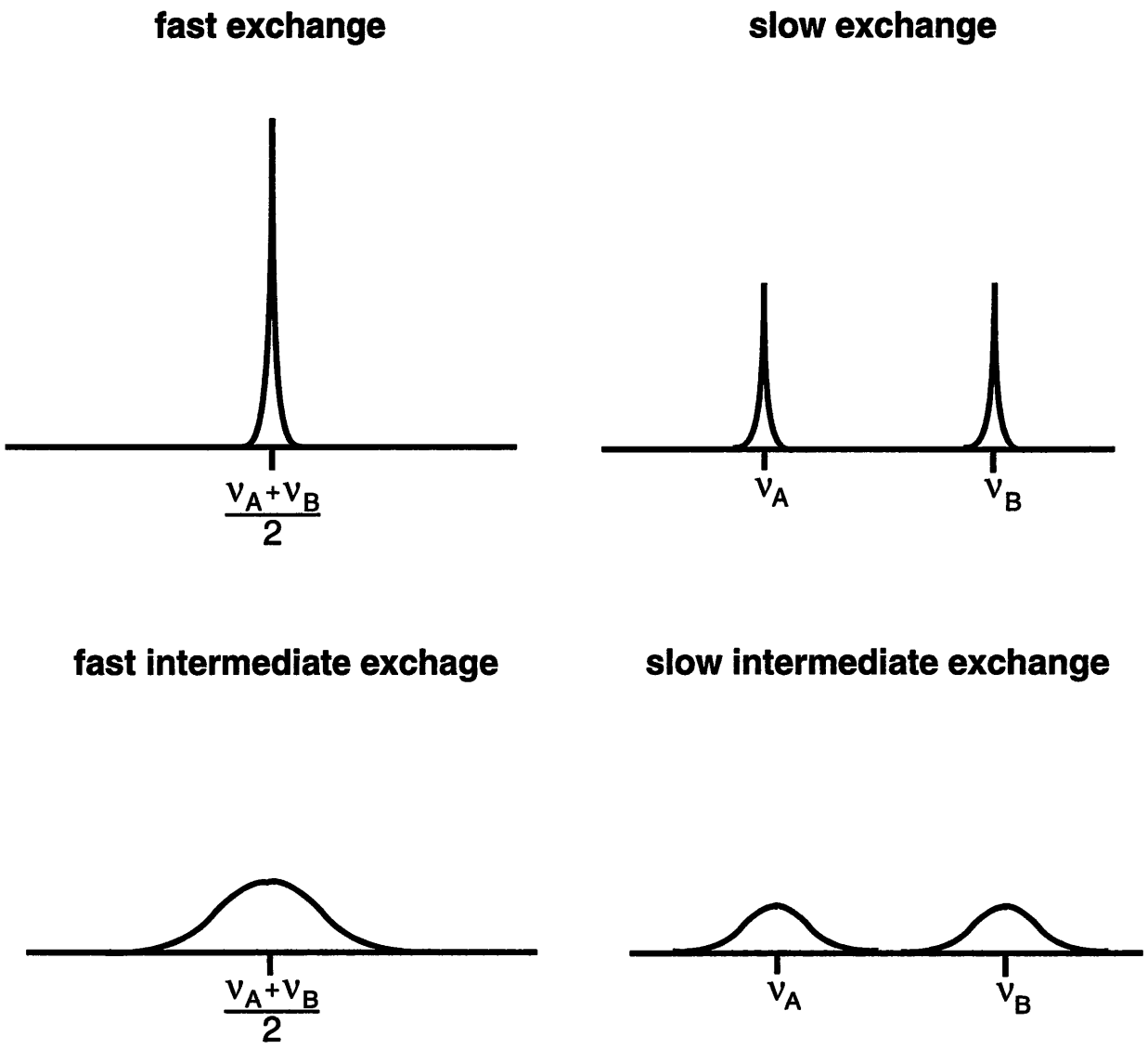


Figure 2.16: The effect of exchange processes on observed linewidths for a proton that occupies states A and B equally, and resonates at frequency ν_A when occupying state A and frequency ν_B when occupying state B. As the rate of exchange moves from the limiting case of fast exchange to the limiting case of slow exchange, there is a kinetic regime of intermediate exchange in which extreme line broadening occurs, a significant problem for NMR.

these two limiting cases is *intermediate exchange* in which $k_{AB} \cong 2\pi(\Delta\nu_{AB})$ and is characterized by extreme line broadening of resonances as they move between the two limiting cases. Processes on the 1-50 millisecond timescale typically result in intermediate exchange.

The T_2 (transverse) relaxation time, which describes the rate of decay of the observable magnetization after a pulse, is the primary determinant in the overall linewidths observed for resonances for a given macromolecule (Bax & Grzesiek, 1993). For larger macromolecules, shortening of the T_2 relaxation time comes primarily from an increase in the molecular correlation time, τ_c (which also impacts the nuclear Overhauser effect, as described in equation 8), that describes the rate at which a macromolecule tumbles in solution (Wüthrich, 1986). This gives rise to the infamous ~30 kDa limitation for the size of macromolecules that can be feasibly studied using NMR spectroscopy. The relation

$$\tau_c \propto \frac{\mu}{kT} \qquad (9)$$

in which μ is the intrinsic viscosity, k is Plank's constant, and T is the temperature, implies that the relaxation rates can be influenced by the temperature at which spectra are taken and the concentration of the macromolecule in solution (which affects the solution viscosity). To solve the structure of larger proteins and RNA, a common tactic has been to label with ^{13}C and ^{15}N to take advantage of 3D- and 4D-NMR experiments which relieve the spectral crowding problem associated with these molecules. However, the incorporation of these isotopes into macromolecules creates efficient heteronuclear dipolar relaxation pathways

which further shorten the T_2, yielding broader linewidths (Wagner, 1993). As many of the interesting biological macromolecules are in this molecular weight region, it has become the focus of much study in the NMR field to develop new techniques that ameliorate many of the problems associated with high molecular weight.

Sample preparation

The conditions for the synthesis of RNA transcripts using *in vitro* T7 RNA polymerase run-off transcription is detailed in Appendix 1, and the synthesis of precursor ribonucleotide triphosphates labeled with $^{2}H/^{13}C/^{15}N$ is described in Appendix 2. Typically, 45-60 ml of transcription reaction were necessary to obtain sufficient RNA to make a 600 µl NMR sample containing 1 mM RNA. Transcription reactions were set up in 15 ml aliquots containing all of the reaction components and incubated in the warm room (37 °C) for 4 hours. The reactions were quenched by the addition of an equal volume of phenol (Ameresco, buffered against TE, pH 8.0), vortexing vigorously to extract the protein, and centrifuging in a JA-17 rotor (Beckman) at 5000 rpm for 5 minutes to separate the phases. The aqueous layer was removed, and the phenolic layer back-extracted with an additional 5 ml ddH_2O to further recover any RNA. The aqueous phases were combined for each reaction to yield 20 ml solution, to which 2 ml 3.0 M sodium acetate, pH 5.2 and 50 ml 100% ethanol was added to precipitate the RNA. The precipitate was pelleted by centrifugation in a JA-17 rotor at 15,000 rpm for 30 minutes, dried in a Speed-Vac for five minutes after the supernatant was decanted off, and resuspended in 1 ml ddH_2O. To the RNA from each 15 ml aliquot (now resuspended in 1 ml ddH_2O), 1 ml Formamide stop buffer was added, and loaded onto a preparative (45x35x0.4 cm) 12% (29:1) polyacrylamide gel containing 8 M urea and 1xTBE running buffer. The gels were run at 60 watts, constant wattage, for ~14 hours, until the xylene cyanol approached the bottom

of the gel. The RNA was visualized by UV shadowing on a fluorescent thin layer chromatography plate, excised, and electroeluted from the gel using an Elutrap (Schleicher & Schuell) at 4 °C at 200 V, constant voltage, until the gel slices appeared transparent to UV light. The eluted RNA was ethanol precipitated and resuspended in 1 ml of TE buffer, and dialyzed in a microdialysis chamber fitted with 3500 molecular weight cutoff dialysis membrane (cellulose ester, Spectrum), and dialyzed at 4 °C against a series of solutions in the following order (Wyatt *et al.*, 1991):

(1) 10 mM sodium phosphate, pH 6.4, 10 mM EDTA, 100 mM sodium chloride

(2) 10 mM sodium phosphate, pH 6.4, 0.1 mM EDTA, 100 mM sodium chloride

(3) ddH_2O.

One liter of each solution, recirculated using a peristaltic pump, was dialyzed against the RNA sample (along with sufficient protein BS15, purified as described in Appendix 1, to form a 1:1 complex with the RNA, in an adjacent well in the microdialysis chamber) for 12 hours, and the dialysis against deionized water was repeated twice. These samples were lyophilized to dryness in a Speed-Vac, resuspended in 600 µl of the appropriate NMR buffer and relyophilized. For NMR experiments in H_2O, the lyophilized RNA was resuspended in 540 µl H_2O and 60 µl D_2O, a few specks of solid trimethyl-silyl-propionic acid (TSP, Sigma) was added, and placed in a Wilmad 535-PP NMR tube. For experiments in D_2O, the lyophilized sample was resuspended in 500 µl 99.9% D_2O (Cambridge Isotope Laboratories), relyophilized to dryness, and repeated three times. After the final exchange, the sample was resuspended in 600 µl 99.996% D_2O (Isotec) under argon, and placed in a dry Wilmad 535-PP tube with a speck of TSP to serve as a reference.

NMR spectroscopy

The NMR spectra discussed in this thesis were acquired on several different instruments. One dimensional spectra at were acquired on a Varian VXR-500 MHz instrument using a 5 millimeter Varian inverse detection probe at the M.I.T. Department of Chemistry Spectroscopy Lab. Experiments performed in a buffer containing 90% H_2O/10% D_2O to observe the exchangeable proton spectrum were recorded at 25 °C using a 1331 binomial solvent suppression (Hore, 1983) with a sweep width of 12000 Hz, 16K complex data points, and 128 scans (unless stated otherwise in a figure legend). The excitation maximum was set at 3500 Hz, at the imino proton region of the spectrum. 2D-NOESY and DQF-COSY spectra were acquired on the same instrument at 45 °C in a buffer containing 99.996% D_2O (Isotec). Data sets with 2048 complex points in the t_2 dimension and 512 complex points in the t_1 dimension with 32 scans per slice and a sweep width of 5500 Hz in each dimension were acquired. NOESY spectra were acquired using 50 and 400 millisecond mixing times. An additional NOESY spectrum was taken on a Varian UNITY 750-MHz instrument at the Francis Bitter National Magnet Laboratory using a mixing time of 200 milliseconds, with 8192 t_2 complex data points, 1024 t_1 data complex points with a sweep width of 8000 Hz, and 32 scans per slice at 45 °C. Heteronuclear 2D-HSQC (constant time) and HMQC spectra of ^{13}C-labeled RNA (Santoro & King, 1992) were acquired on a 500 MHz instrument built at the Francis Bitter National Magnet Laboratory. Heteronuclear 2D spectra with 2048 complex data points and a sweep width of 6000 Hz in the t_2 dimension and 128 complex data points and a sweep width of 5000 in the t_1 dimension were acquired at 45 °C. The ^{13}C transmitter frequency was set to 19000 Hz, with GARP decoupling during acquisition. Data were processed using NMR Pipe software (developed by NIH & Laboratory of Chemical Physics, NIDDK) on Silicon

Graphic workstations, with a combination of exponential and gaussian weighting functions.

Chapter 3: Determination of the minimal rRNA binding site for S15

The following studies of the S15-rRNA interaction employ a variant from the ribosome of the moderate thermophile *B. stearothermophilus*, a system that has been utilized in a number of other ribosomal structural studies. *Bacillus stearothermophilus* S15 (BS15) has nearly 60% identity to the *E. coli* protein (Figure 1.8), and the 16S rRNA in the S15 binding region also shows strong sequence and secondary structure homology, suggesting that these two proteins bind to 16S rRNA in a very similar fashion. It has been demonstrated that functional hybrid 30S ribosomal subunits can be reconstituted by mixing *E. coli* rRNA and *B. stearothermophilus* ribosomal proteins, further suggesting that the ribosomal proteins and 16S rRNA are highly conserved in their structure and function (Nomura *et al.*, 1968; Higo *et al.*, 1973). Starting with a fragment of 16S rRNA comprising nucleotides 585-756 that contains all of the elements implicated by previous studies as important for S15 binding (Figure 1.9), a binding assay to monitor the specific interaction between S15 and rRNA was developed. Then, specific deletions were introduced into the RNA and tested for binding, in order to define the regions of the 16S rRNA that are necessary for S15 recognition, and develop a minimal RNA site suitable for structural characterization of this protein-RNA interaction.

Interaction of BS15 protein with Fr1 RNA

These studies were initiated by cloning the homologous region to nucleotides 585-756 in *E. coli* 16S rRNA (Figure 1.2) from the *B. stearothermophilus* 16S rRNA gene

(rrnB) and inserting into a plasmid containing the T7 RNA polymerase promoter (see Appendix 1 for details on the construction of plasmids containing sequences that were used as templates for the synthesis of the RNAs used in these studies). Additional sequence was added at the 3' end of the RNA fragment gene such that the 5' and 3' ends of the RNA would form a nearly perfectly base paired helix, in order to minimize the potential for alternative RNA conformations induced by the presence of single stranded tails at the 5' and 3' ends of the RNA. A *Sma*I site was placed at the 3' end of the gene for linearizing the plasmid for runoff transcription. This results in a CCC at the 3' end of the transcribed RNA that is complementary to the GGG necessary at the 5' end of the RNA for efficient transcription by T7 RNA polymerase (Milligan *et al.*, 1987). Fr1 RNA should contain all of the elements of specificity for the S15-rRNA interaction based upon previous work performed in the *E. coli* system (Fr1 RNA is illustrated in Figure 3.3). T_1 nuclease mapping indicated that Fr1 RNA adopted a secondary structure similar to that found in the central domain of 16S rRNA.

The interaction of ^{32}P-labeled Fr1 RNA and BS15 protein was monitored using a native gel mobility shift assay. Titration of the BS15 protein into Fr1 RNA at a concentration of approximately 4 nM yields a complex of higher electrophoretic mobility than the unbound RNA, as shown in Figure 3.1. To determine the apparent dissociation constant for this RNA-protein interaction, the amount of ^{32}P present in the free and bound RNA bands was quantitated and the binding data was fit to an equation that describes a simple bimolecular equilibrium, yielding a value of 35±12 nM (Table 3.1). The binding titrations were performed at 25 °C as well as 4 °C, and the apparent dissociation constants for the protein-RNA interaction were the same at both temperatures, within the error of the experiment (Table 3.1). Previous experiments using mixed components from *E. coli* and *B. stearothermophilus* 30S subunits indicated that BS15 could be specifically incorporated into functional 30S subunits in which all other components were derived from *E. coli*

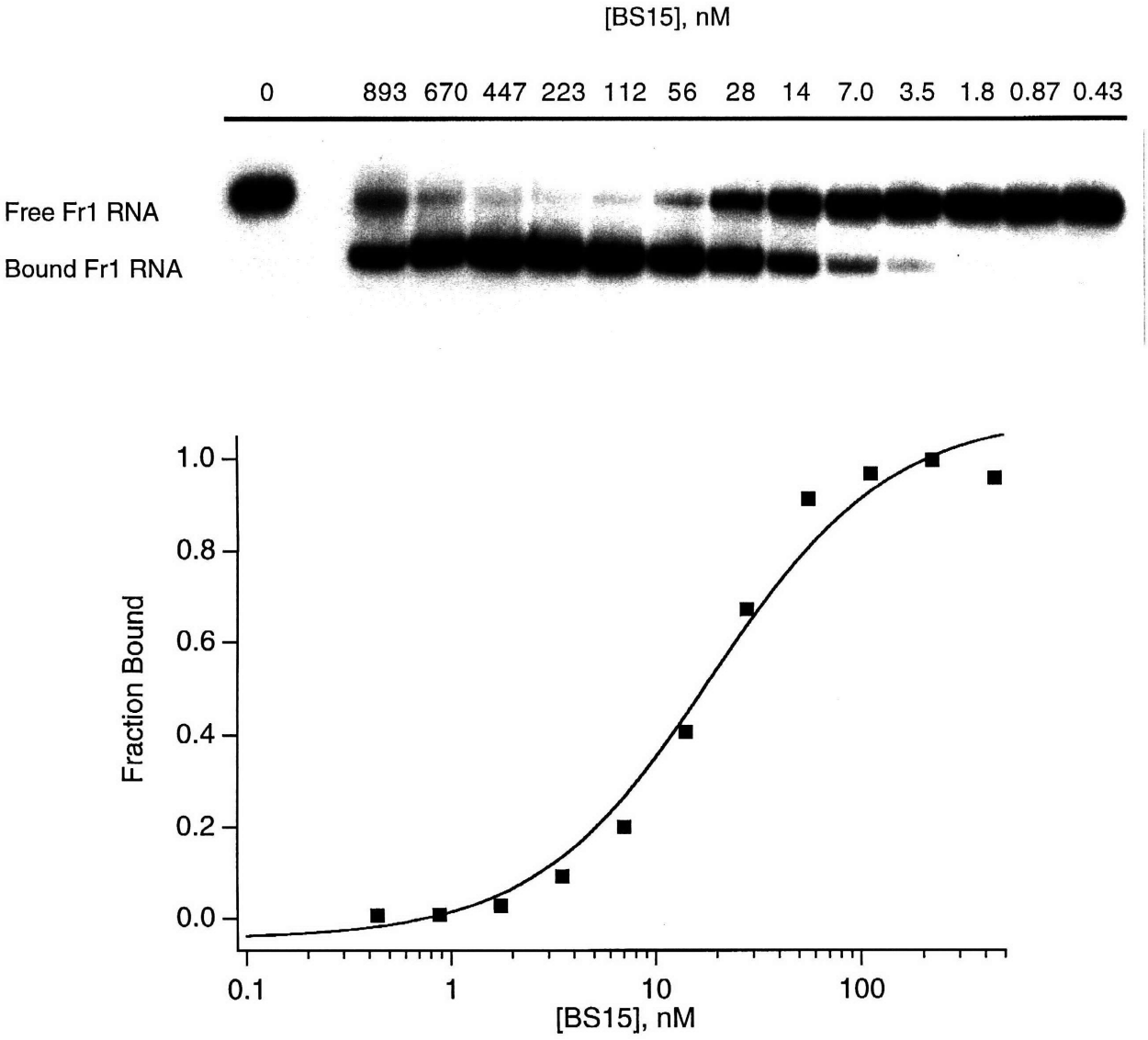


Figure 3.1: Titration of BS15 protein into ^{32}P-labeled Fr1 RNA. This individual titration, performed at 4°C, yields an apparent dissociation constant of approximately 25 nM for the BS15-Fr1 RNA interaction.

Table 3.1: Apparent equilibrium dissociation constants for the BS15-Fr1 RNA interaction, as measured with the native gel mobility shift assay using different biochemical methods.

Method	4 °C	25 °C
Direct titration	35 ± 12 nM	33 ± 13 nM
Self competition	28 ± 12 nM	
Kinetics		35 ± 14 nM

(Higo *et al.*, 1973). To verify that the interaction between BS15 and ES15 and 16S rRNA was conserved, the ability of ES15 to bind Fr1 RNA was tested using the gel shift assay. Like BS15, ES15 was capable of inducing a higher mobility complex upon binding and had an apparent dissociation constant of 30±10 nM (data not shown). Thus, the S15-rRNA interaction appears to be conserved between these two evolutionarily distant prokaryotes.

The binding of Fr1 RNA was compared to that of 16S rRNA purified from *Bacillus stearothermophilus* using a competition assay in which increasing concentrations of 16S rRNA was used to compete for binding of BS15 with ^{32}P-labeled Fr1 RNA. The apparent dissociation constant of the *B. stearothermophilus* S15-16S rRNA interaction, 130 nM (Table 3.2), is consistent with values obtained for the *E. coli* variant binding to *E. coli* 16S rRNA (Schwarzbauer & Craven, 1981). BS15 bound with moderately higher affinity to Fr1 RNA than 16S rRNA, demonstrating that all of the elements of the specific BS15-16S rRNA interaction were preserved in Fr1 RNA.

The observed binding titration between BS15 and Fr1 RNA was consistently steeper than the fit of the data to an equation describing single ligand binding, suggesting the potential for multiple proteins binding to Fr1 RNA. At protein concentrations higher than 400 nM, a second band of lower electrophoretic mobility appeared (Figure 3.1), further suggesting that BS15 could bind to multiple sites within Fr1. To determine the stoichiometry of the higher mobility complex, the protein was titrated into 260 nM Fr1, a concentration well above the BS15-Fr1 apparent dissociation constant. As shown in Figure 3.2, the protein saturated the RNA at 290 nM BS15, demonstrating that the higher mobility species was a 1:1 complex between BS15 and Fr1. Above a stoichiometry of 1.5:1 BS15:Fr1, lower mobility species appeared, suggesting that these species corresponded to multiple proteins binding Fr1 RNA, presumably in a non-specific fashion.

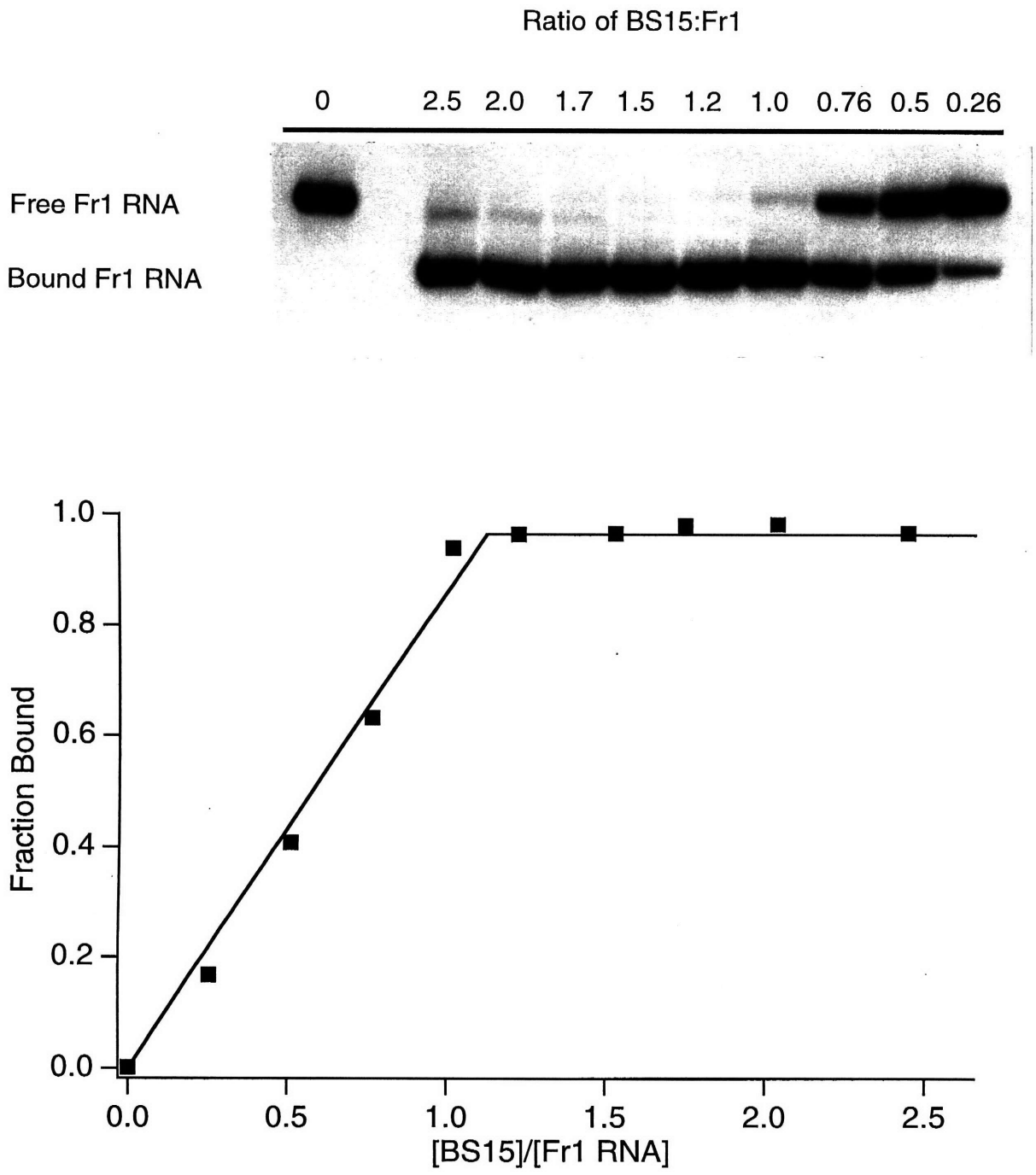


Figure 3.2: Titration of BS15 protein into reactions containing 260 nM ^{32}P-labeled Fr1 RNA. The breakpoint of the titration occurs at a ratio of 1.1:1 BS15:Fr1 RNA, indicating that the high mobility complex has a 1:1 protein:RNA stoichiometry.

Systematic deletion of Fr1 RNA to produce a minimal binding site for BS15

As a prerequisite for structural studies of this protein-RNA interaction, it was necessary to develop a minimal RNA site for specific protein binding. This was accomplished by systematic deletion of entire helices or portions of helices from Fr1 RNA using recombinant PCR methods (Figure 3.3). Each deletion was capped with a stable UNYG or GNRA type tetraloop to minimize the potential for alternative secondary structure formation as a result of the deletion (Molinaro & Tinoco, 1995). Initially, the ability of the deletion mutants to interact with BS15 was tested using the direct titration method. However, as illustrated in Figure 3.4, the protein-RNA complexes demonstrated markedly different mobility shift patterns relative to free RNA for each deletion mutant. While Fr1 and Fr4 RNA showed a mobility shift enhancement upon protein binding, Fr2 RNA showed a slight mobility retardation, and Fr3 and Fr7 RNAs showed no apparent binding. The variable behavior of the mobility of the protein-RNA complex made the interpretation of the gel shift assay difficult. Although a protein's inability to shift an RNA probe in the gel matrix upon titration of protein is generally interpreted as lack of binding activity, considering that both mobility enhancement and retardation is observed with this system, it was conceivable that one of these RNAs bound protein with high affinity, but that the free and bound RNA probe comigrated in the gel. Thus, the ability of BS15 to bind to the deletion mutants was tested using a competition assay in which increasing concentrations of unlabeled mutant RNA was used to compete for binding of BS15 with ^{32}P-labeled Fr1 RNA. The advantage of this assay is that the ability of a mutant RNA to bind S15 was monitored by observing the same signal, complex formation between Fr1 RNA and BS15, and thus independent of mutant specific effects. Using this assay, one of the deletion mutants that showed no shift in the direct binding assay, Fr7, was clearly able

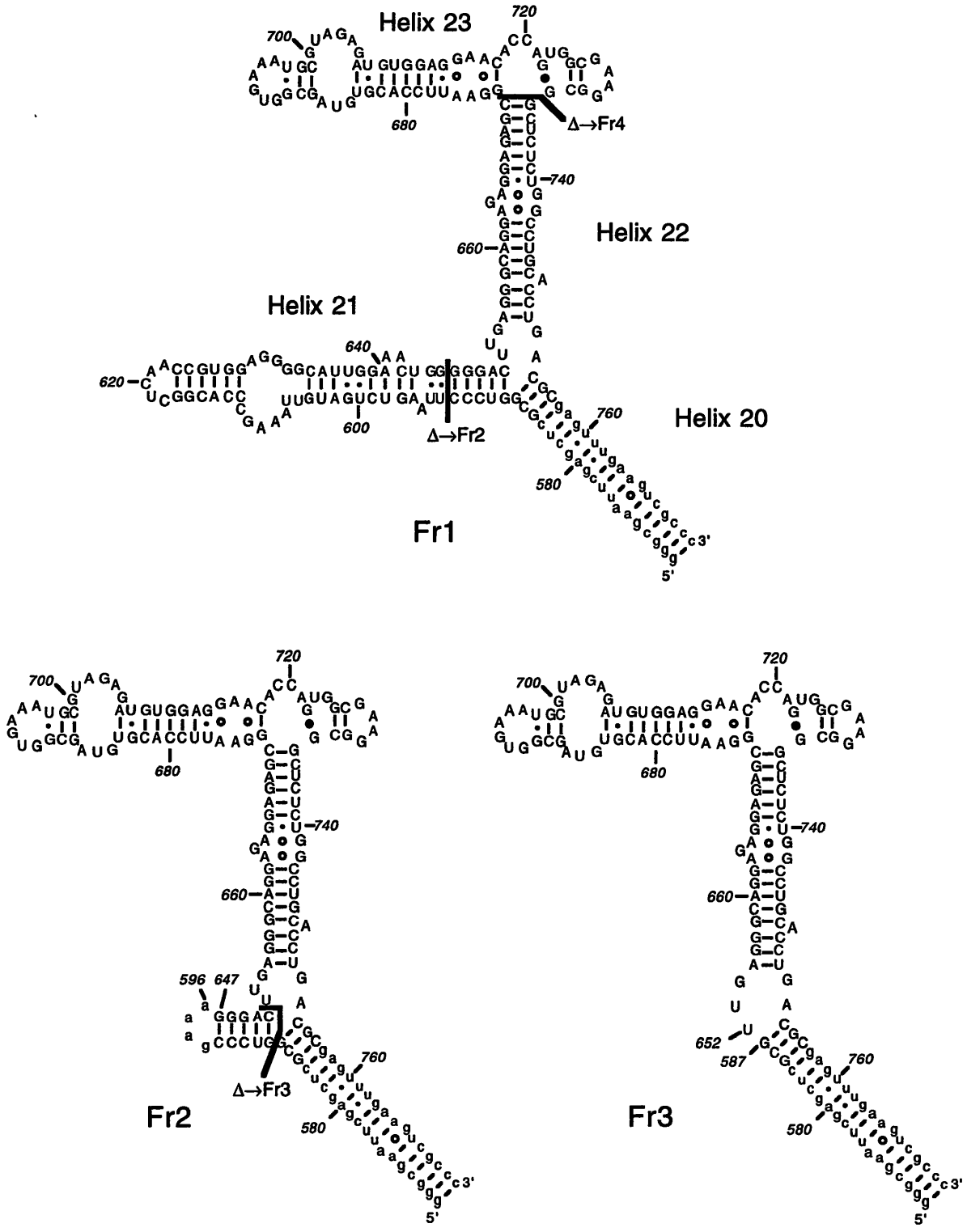


Figure 3.3: Secondary structure of the Fr1 RNA derived from *Bacillus stearothermophilus* 16S rRNA and the family of systematic deletion mutants used in this study. The numbering of the nucleotides follows the same convention as that used for *E. coli* 16S rRNA, and the lower case letters represent non-native sequence. The bold lines denote where the stepwise deletions were made to ultimately yield a minimal RNA substrate for BS15.

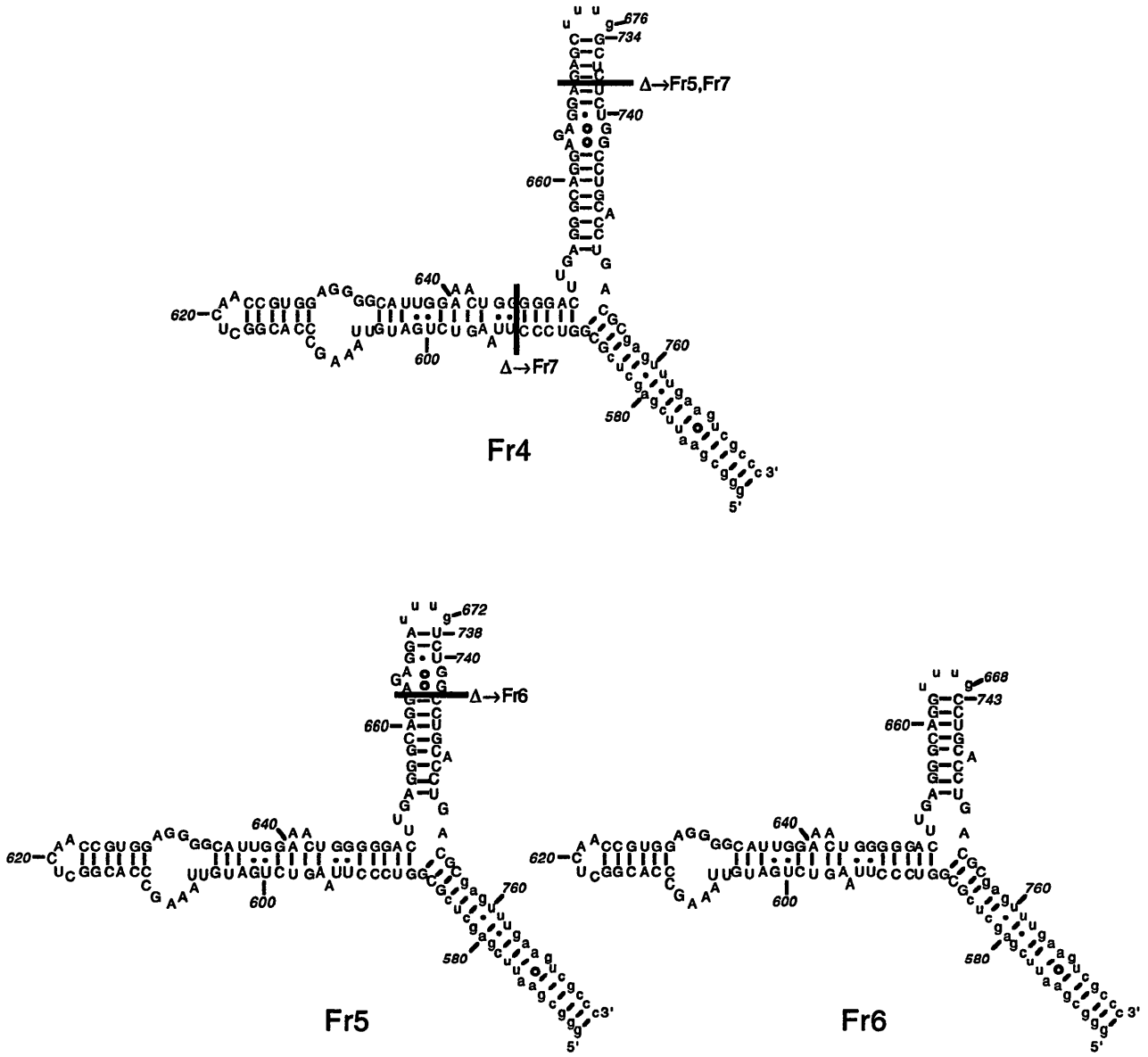


Figure 3.3: Continued.

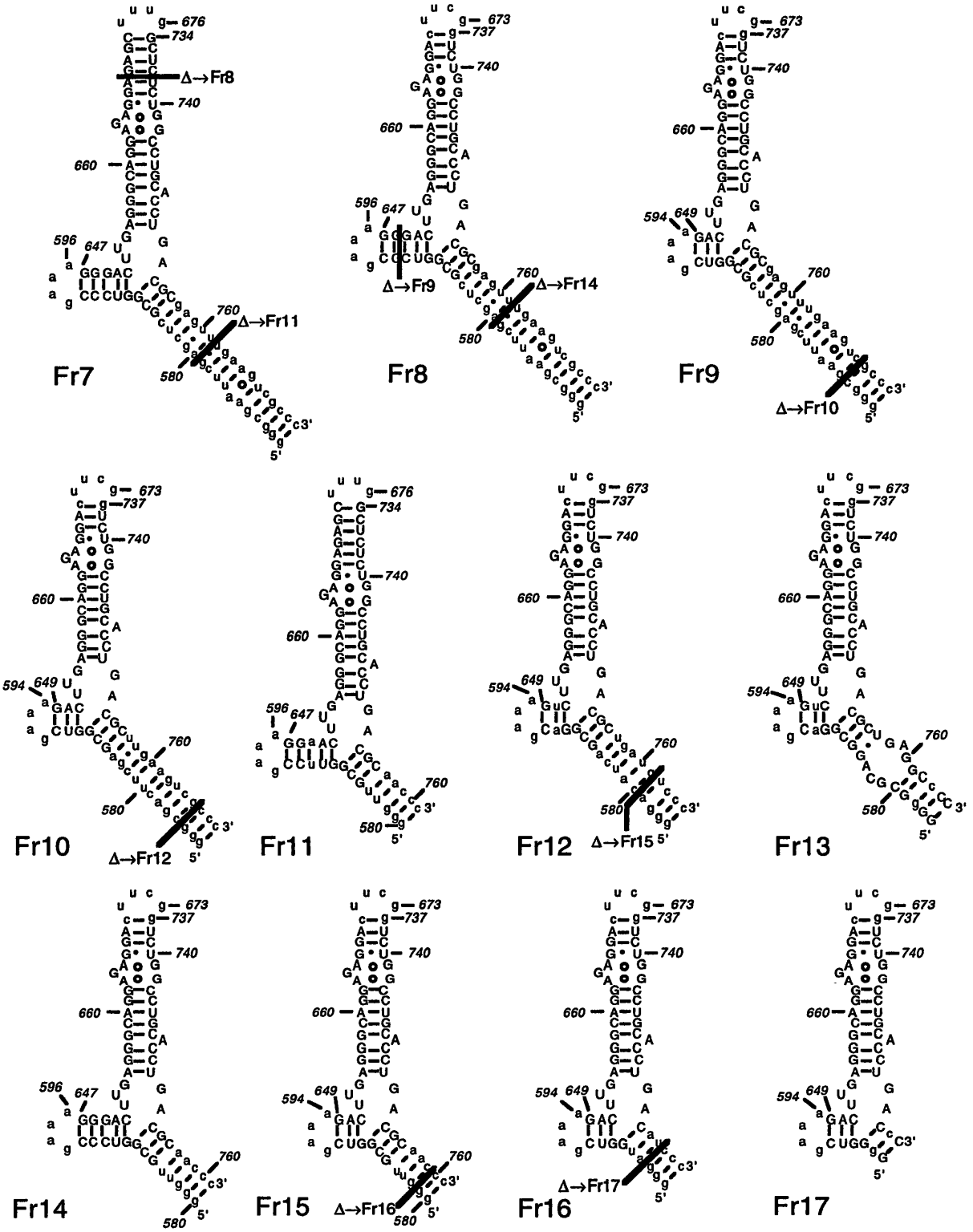


Figure 3.3: Continued.

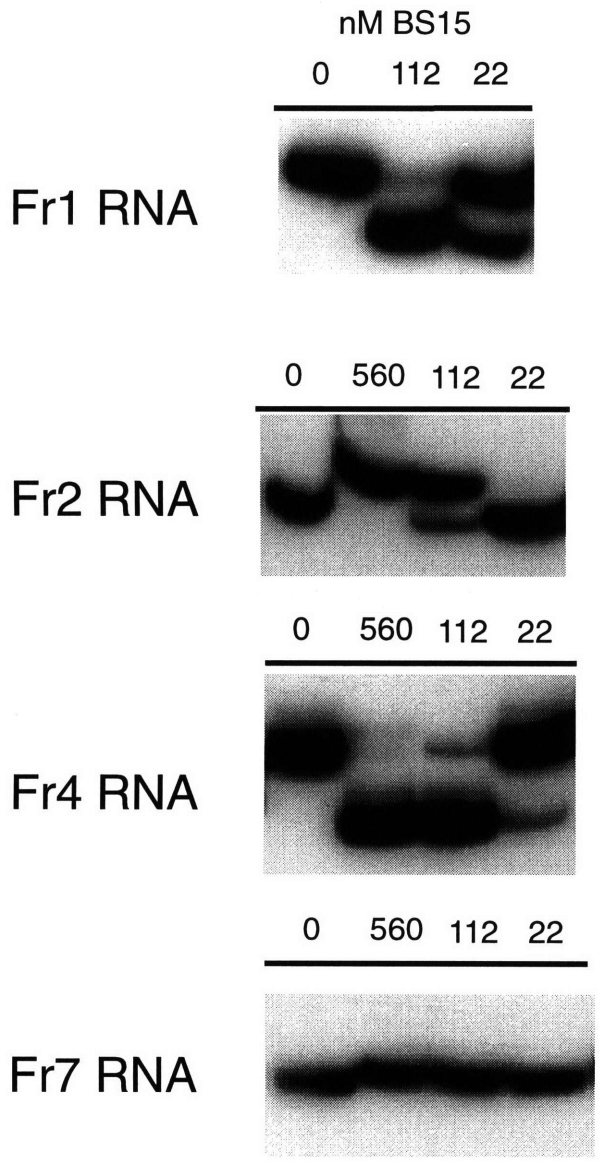


Figure 3.4: Migration of the BS15:RNA complexes relative to nacent RNA for Fr1 and representative deletion RNAs in a 10%, 29:1 polyacrylamide gel matrix. The binding of BS15 to Fr1 and Fr4 RNAs causes the labeled probe RNA to migrate ahead of the unbound protein, while the mobility of bound Fr2 RNA is lower relative to unbound RNA, and binding of BS15 to Fr7 has no affect on the RNAs mobility. All of these RNAs have been shown to bind BS15 with equal affinity using a competition assay.

to bind to BS15 with wild type affinity. This was later confirmed by direct titration of BS15 into ^{32}P-labeled Fr7 using a more highly crosslinked gel matrix. Representative competition experiments are shown in Figure 3.5, and the results are summarized in Table 3.2.

Helix 21 of 16S rRNA has previously been identified as the binding site of the primary binding protein S8 (Mougel *et al.*, 1987; Svensson *et al.*, 1988), and the results suggest that most or all of helix 21 is dispensable for S15 binding. Fr2 RNA, in which all but the first five base pairs of helix 21 are deleted, bound to BS15 with wild type affinity. Complete deletion of helix 21, in Fr3 RNA, converted the three-way junction between helices 20, 21 and 22 into an internal loop. This RNA does not bind BS15 with high affinity, having a dissociation constant that is consistent with non-specific binding. Other RNAs where helix 21 is truncated to three base pairs (Fr9, 10, 12, 13, 15-17) also bind BS15 with wild type affinity. In Fr11, the identity of the third base pair from the junction is altered from a C-G to a A-U pair, and in Fr12 and Fr13, the second base pair from the junction is altered from a U-A to a A-U pair. These RNAs bind BS15 with wild type affinity, demonstrating that the identity of these base pairs in helix 21 are not important for recognition. Thus, this set of deletion RNAs demonstrate that part of helix 21 is necessary for maintaining the integrity of the junction, but that the identity of the base pairs within helix 21 are not important for S15 binding.

Chemical footprinting experiments implicate the highly conserved tetraloop in helix 23 and the conserved purine rich internal loop as the region of 16S rRNA contacted by S15 (Mougel *et al.*, 1988; Svensson *et al.*, 1988). However, the complete deletion of helix 23 in Fr4 RNA does not lead to an appreciable loss of binding affinity, demonstrating that the tetraloop is not an important determinant in the specificity of the S15-rRNA interaction. The distal four base pairs of helix 22 (nucleotides 669-672/734-737) are also nonessential for binding, as demonstrated by high affinity binding to Fr5. Deletion of the internal loop, which contains several phylogenetically conserved purine-purine base pairs along with a

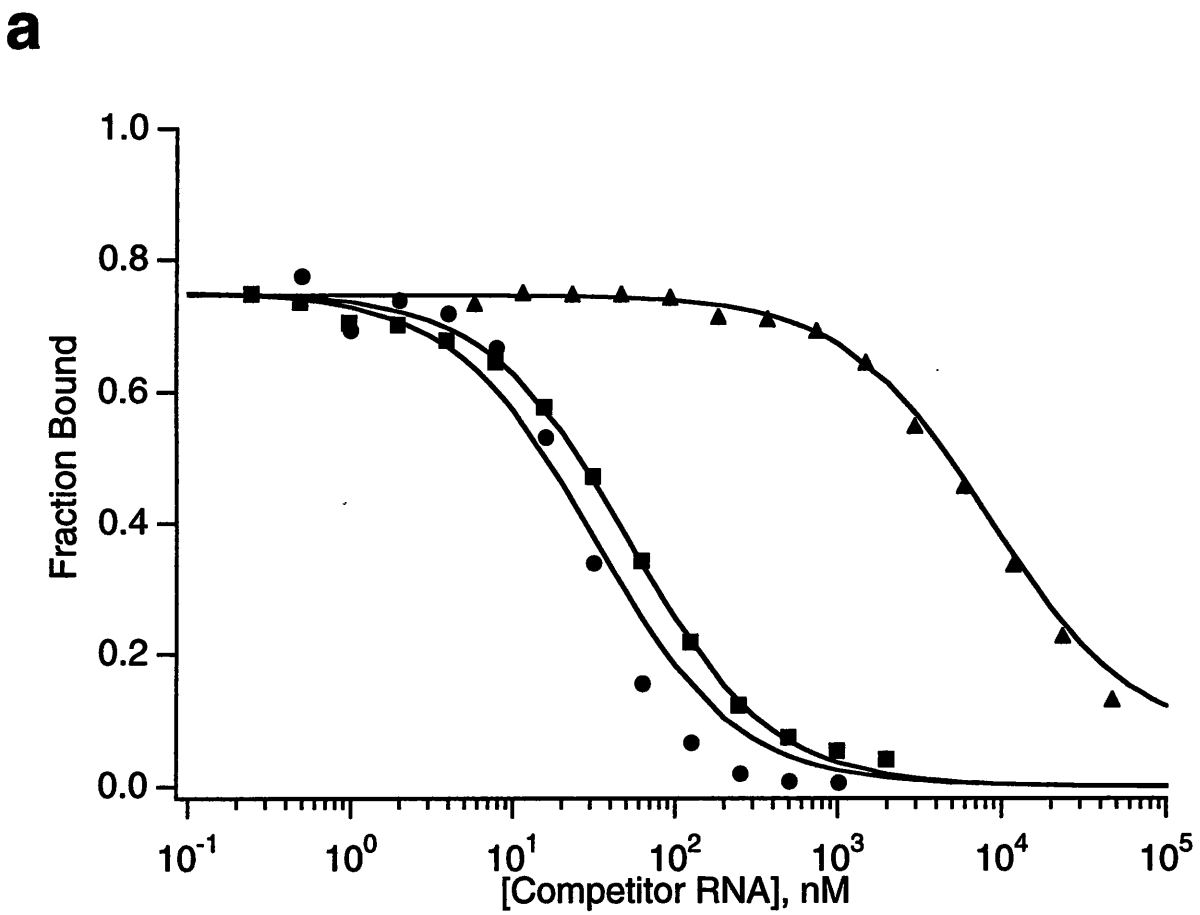


Figure 3.5: Representative native gel competition assays in which unlabeled competitor RNA was used as a competitor of a BS15-^{32}P-labeled Fr1 RNA complex. (a) Graphical representation of the data for these experiments with competitor RNAs Fr1(■), Fr2 (●), and Fr3 (▲). (next page, b-d) Autoradiograms of competition experiments using Fr1, Fr2, and Fr3 RNAs are shown in a, b, and c, respectively.

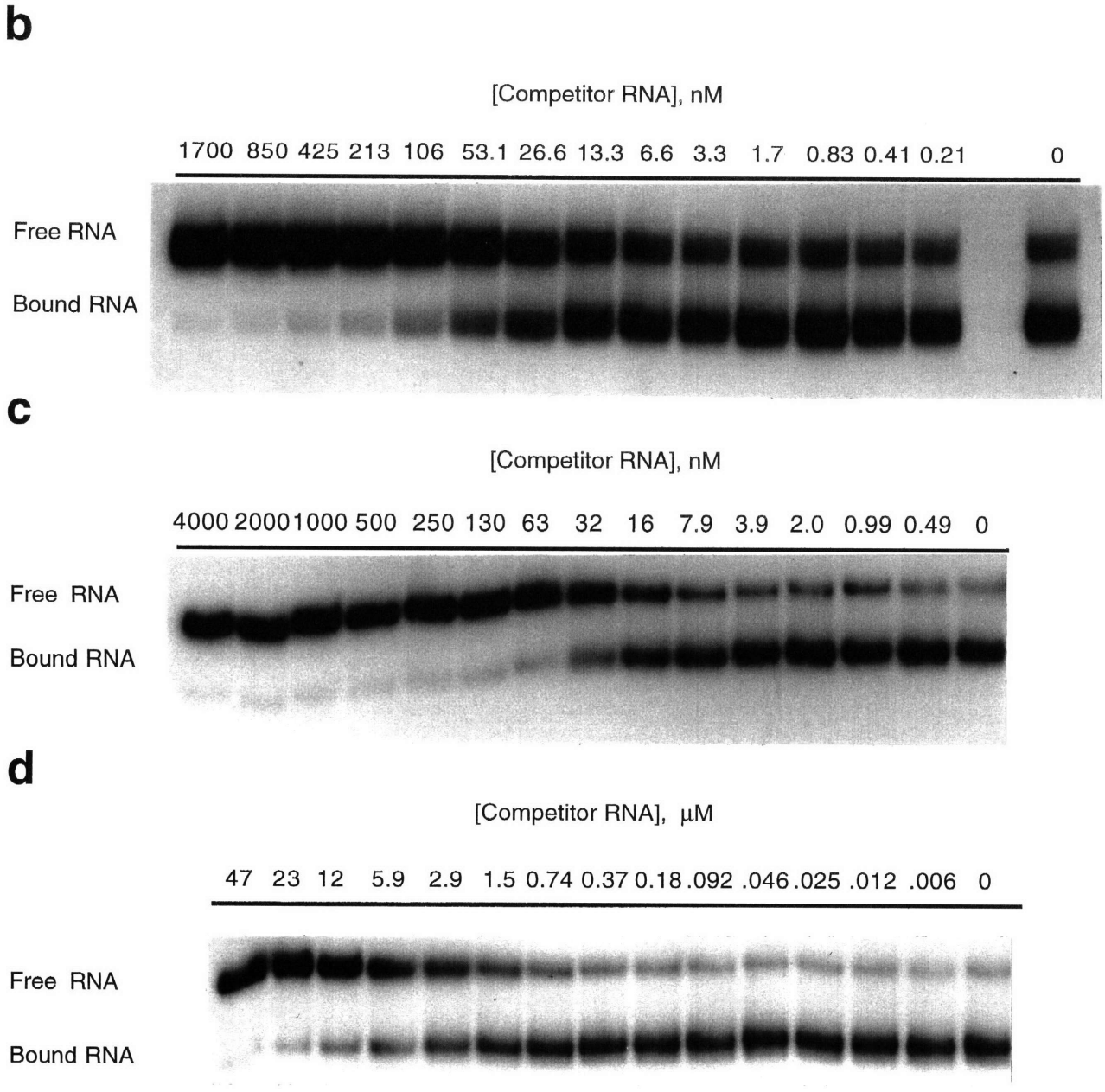


Figure 3.5: Continued.

Table 3.2: Relative apparent dissociation constants as measured by competition assays and dissociation and association rates of select RNA fragments†.

16S rRNA Fragment	K_{rel}*, 4 °C	k_{off} (s^{-1}), 25 °C	k_{on}($M^{-1}s^{-1}$), 25 °C
16S rRNA	3.7 3.8		
Fr1	1.0	6.70×10^{-4}	1.91×10^{4}
Fr2	0.88 0.70		
Fr3	180		
Fr4	0.44	2.85×10^{-4}	
	0.37	2.80×10^{-4}	
Fr5	0.85 0.63		
Fr6	23 20		
Fr7	0.44	2.63×10^{-4}	
	0.48		
Fr8	1.0	2.58×10^{-4}	
	0.83		
Fr9	0.87 0.65		
Fr10	0.43		
Fr11	0.55		
Fr12	0.43		
Fr13	0.59		
Fr14	0.54		
Fr15	0.48	2.40×10^{-4}	1.10×10^{-4}
Fr16	0.65		
Fr17	85		

† All individual measurements of K_{rel} by competition are listed. All deletion mutants were also tested by direct titration to verify these results (data not shown).

* $K_{rel} = (K_{d,competitor}/K_{d,Fr1})$

highly conserved bulged guanosine, in Fr6 gives rise to more than a tenfold decrease in binding affinity. This clearly implicates helix 22 as important for the high affinity binding of BS15 to 16S rRNA as had been suggested earlier (Müller *et al.*, 1979; Zimmerman & Singh-Bergmann, 1979; Mougel *et al.*, 1988; Svensson *et al.*, 1988).

In helix 20 of Fr1 RNA, only the three base pairs closest to the three way junction between helices 20, 21 and 22 have the same identity as those in the 16S rRNA. To test the importance of the identity of these base pairs, the wild type helix 20 except for one base pair was created in Fr13. This RNA has the same binding affinity as Fr12 RNA, demonstrating that the sequence of this helix is not important for the specific BS15-RNA interaction. The importance of the length of helix 20 for the high affinity RNA-protein interaction was tested using a series of systematic deletions generating helices ranging from 3 to 19 base pairs in length. RNA fragments containing a helix 20 of length 6 to 19 base pairs in length do not have significant differences in their binding affinities. Also, in Fr16, the only wild type base pair is the G-C base pair closest to the three-way junction, suggesting that this is the only base pair which might be important for S15 recognition. Fr17 RNA, which contains a three base pair helix 20, demonstrates a sharp drop in the binding affinity of BS15. Like helix 21, only a small portion of helix 20 is necessary to maintain the integrity of the junction and high affinity binding, and in each, only the identity of the most apical base pair to the three way junction may be necessary for the specific interaction.

Kinetics of the BS15-RNA interaction

In order for the minimal RNA substrates to serve as suitable models of the wild type RNA-protein interaction for structural studies, they must demonstrate similar kinetic as well as thermodynamic properties (Long & Crothers, 1995). The dissociation rates of various RNA fragments were determined using a native gel mobility assay in which a pre-

formed complex was challenged with 50-200 nM unlabeled competitor RNA, and the amount of complex remaining was monitored over time on a native polyacrylamide gel (Figure 3.6). Representative RNAs were chosen at various points during the process of minimization of the RNA site, and the derived values of k_{off} are given in Table 3.2. As a control for the possibility of facilitated dissociation by the competitor RNA, the dissociation rate of Fr1 was measured using a dilution protocol in which the pre-formed complex was diluted 100-fold to monitor the dissociation reaction. For Fr1 RNA, both dissociation by competition and by dilution give values of approximately $6.7x10^{-4}$ s^{-1} for the unimolecular dissociation rate. Furthermore, for all RNA fragments tested, the dissociation displays monophasic behavior, indicative of a unimolecular dissociation reaction. This, along with similar kinetic stabilities for all RNAs tested (Table 3.2), provides additional evidence that the nature of the complex has not been significantly affected in the minimal RNA fragments.

The dissociation kinetics of a macromolecular complex strongly influence the quality of NMR spectra due to the influence of chemical exchange, as discussed previously. For exchange processes that occur on the 1-50 millisecond timescale, the linewidths of peaks broaden dramatically, indicative of the "intermediate exchange" problem. To determine whether the S15-rRNA complex would potentially suffer this problem, the half-life of two minimal RNA-protein complexes was determined over the temperature range that would be utilized for subsequent NMR studies. These data are shown in Table 3.3. Even at 45 °C, the half life of this protein-RNA complex is well over 1 minute, demonstrating that even at elevated temperatures, the free and bound states of the RNA will be in slow exchange on the NMR timescale, and thus suitable for detailed structural study by this criterion.

The association rates of Fr1 and Fr15 RNA fragments and BS15 were also measured with the gel mobility shift assay at various protein concentrations. The pseudo-

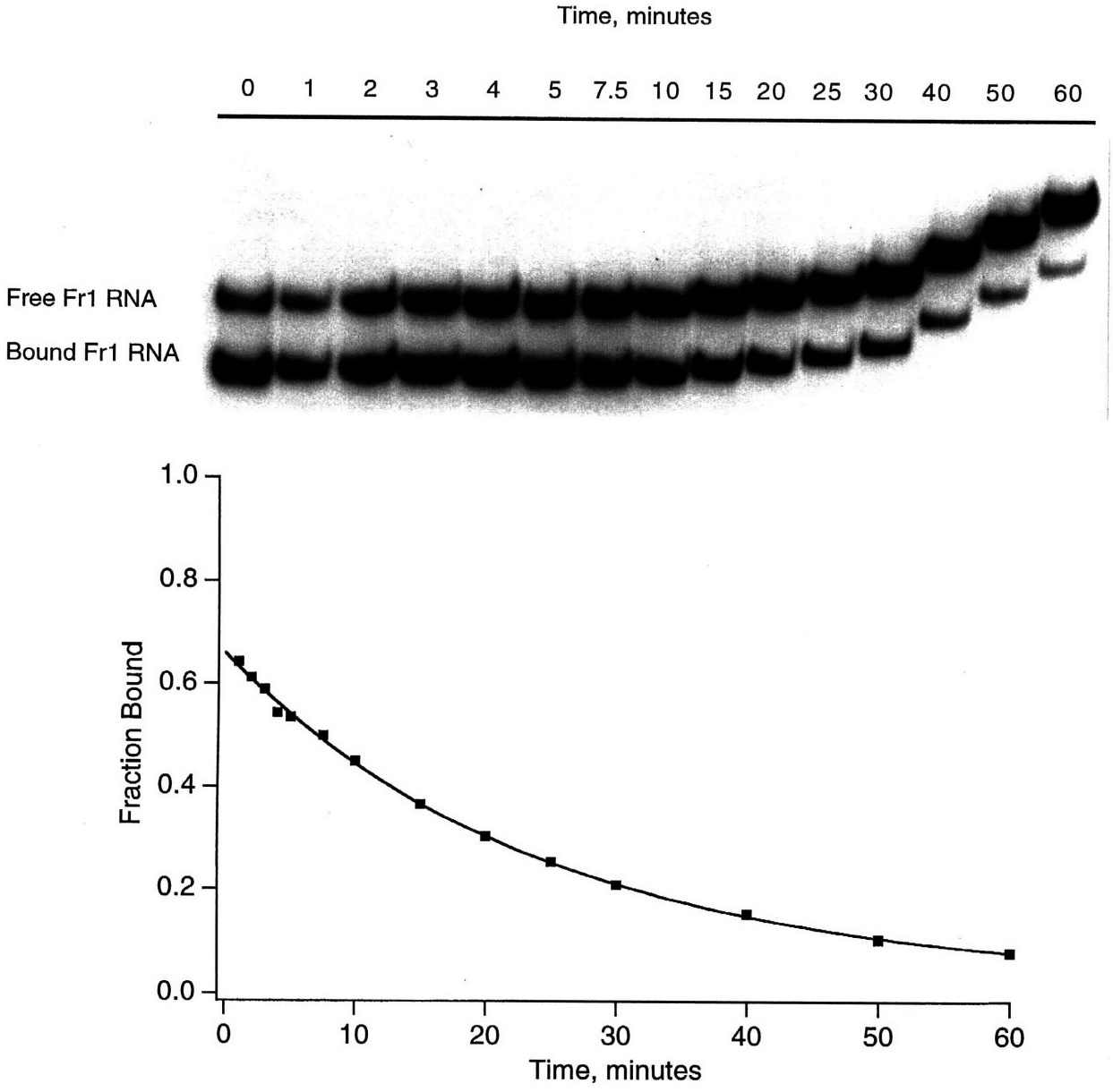


Figure 3.6: Kinetic dissociation of the BS15-Fr1 complex using 50 nM Fr15 RNA to challenge the complex at t=0. This experiment yields a k_{off} of 6.5 x $10^{-4}s^{-1}$ at 25°C, corresponding to a half-life of 18 minutes for the complex.

Table 3.3: Temperature dependence of the dissociation kinetics for the BS15-Fr11 and Fr14 RNAs.

	Fr11 RNA		Fr12 RNA	
Temperature	k_{off}, s^{-1}	$t_{1/2}$, min	k_{off}, s^{-1}	$t_{1/2}$, min
4 °C	2.9×10^{-5}	395	3.1×10^{-5}	371
25 °C	1.7×10^{-4}	68	1.9×10^{-4}	61
45 °C	1.7×10^{-3}	6.9	1.2×10^{-3}	9.8

first order association rate constants show a linear dependence on the protein concentration, demonstrating that this step is bimolecular with an apparent bimolecular rate constant of $1.2x10^4$ $M^{-1}s^{-1}$ for Fr1 RNA (Figure 3.7). An apparent equilibrium dissociation constant for the Fr1 RNA-BS15 interaction at 25 °C can be calculated by dividing the dissociation rate by the bimolecular association rate to yield a value of 35 nM (Table 3.1), in good agreement with the figures obtained by direct titration and self-competition methods. Similarly, the apparent equilibrium dissociation constant measured by this method for Fr15 RNA is consistent with the competition assay. Therefore, the direct titrations and competition assays are able to accurately measure apparent equilibrium dissociation constants for the S15-rRNA interaction.

Discussion

Previous studies have indicated that *E. coli* S15 binds to a fairly limited region within the central domain of the 16S rRNA. However, taken together, these studies provide ambiguous clues as to which portions of this region are most critical for the S15-rRNA interaction. Using a sequential deletion strategy, it has been demonstrated that the binding site for *B. stearothermophilus* S15 in 16S rRNA is centered around the junction between three helices. The conversion of the three-way junction to an internal loop in Fr3 or shortening helix 20 to three base pairs in Fr17 completely abolishes high affinity binding. A stable helix of at least four base pairs is required by S15 in helix 20 only to preserve the integrity of the junction, since the identities of these base pairs, except for possibly the pair proximal to the junction, are not important for S15 recognition. The other region important for S15 binding is the internal loop in helix 22. Deletion of the internal loop in Fr6 diminishes affinity, although not as strongly as deletions near the junction. This suggests that the junction is the primary center of S15 interaction with rRNA, with

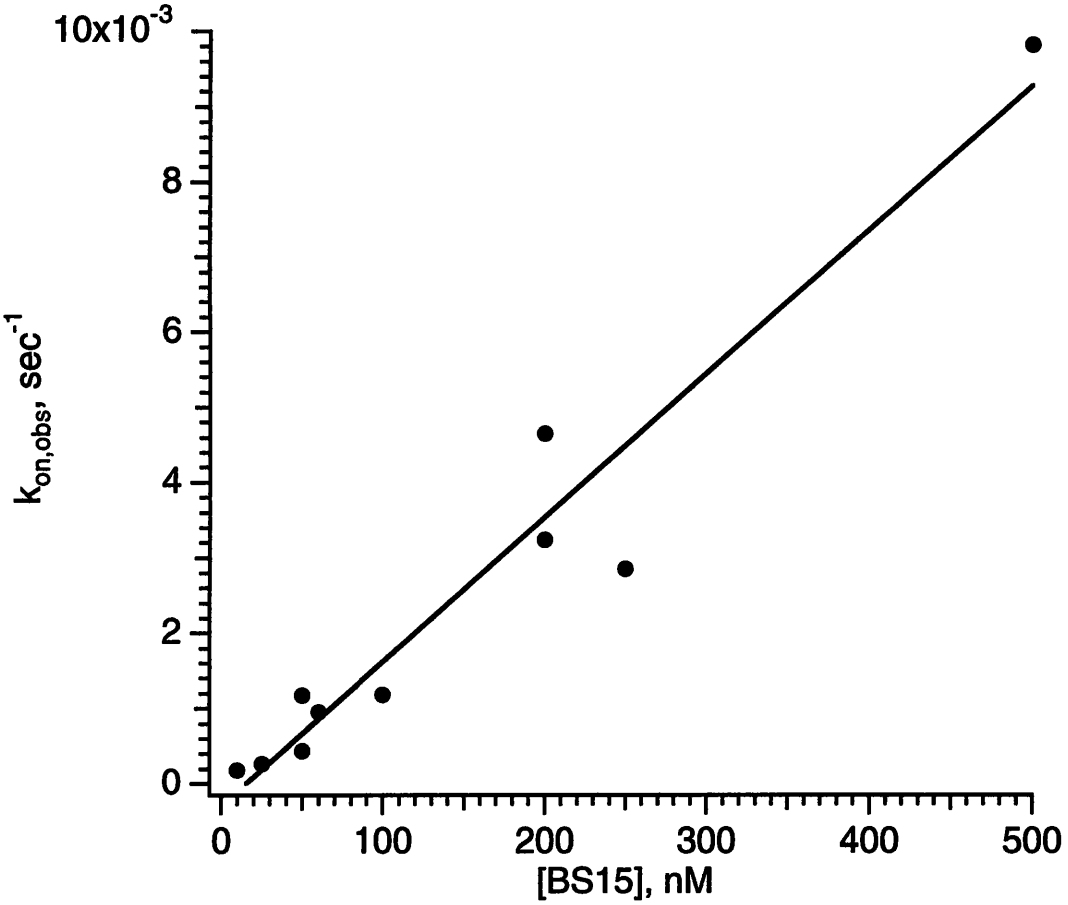


Figure 3.7: The pseudo-first order association rate constant, $k_{on'}$ as a function of protein concentration. The slope of the line yields the bimolecular association rate constant, k_{on}.

other weaker interactions centered around the internal loop of helix 22. These data strongly support the Fe^{2+}-EDTA footprint of S15 bound to 16S rRNA (Powers & Noller, 1995), while suggesting that elements within helix 23 which are protected by S15 using other chemical and enzymatic probes do not represent direct RNA protections by S15.

There are several assumptions that were made in the competition assay which may not hold true. First, in the deletion approach to minimization of the S15 binding site, it was assumed that a loss of a protein-RNA contact would result in a loss of binding activity. However, the apparent equilibrium dissociation constant is a measure of the total free energy of binding, ΔG, and thus in some cases, a intermolecular contact could be lost without affecting the observed K_d. For instance, if the enthalpic (ΔH) cost of a loss of a protein-RNA contact is offset by a favorable entropic (ΔS) change, then the observed free energy of binding, and hence, the apparent K_d, will be in the same for the wild type and deletion mutant RNA. Therefore, there is the formal possibility that there are contacts in helix 23 (or elsewhere in the 16S rRNA) made by S15 which were eliminated during the deletion analysis. Second, the slow association and dissociation kinetics of S15 binding at 4 °C prevent the binding reactions used in the competition assay from coming to full equilibrium during the one hour of incubation prior to application to a gel. This was partially addressed by equilibrating the reactions for one hour at 25 °C, a temperature at which the kinetics of association and dissociation are significantly faster. Since Fr1 RNA shows no significant difference in affinity for BS15 at these two temperatures, the binding reaction should have approached equilibrium by the time the reaction was switched to 4 °C, with an hour being sufficient to equilibrate at the lower temperature. The results of the competition assays were confirmed in part by qualitatively checking the ability of each deletion fragment to bind BS15 using the direct titration approach, an assay that does not suffer from this problem. For every mutant, the direct binding assay correlated well with the values obtained from the competition assay.

The apparent binding constant of S15 for rRNA as measured by various methods is fairly typical for a ribosomal protein. *E. coli* L5, L25, and S8 all display binding affinities within this range at 25 °C (Spierer *et al.*, 1978; Mougel *et al.*, 1986). The specificity of BS15 for its specific site over nonspecific sites is at least 180-fold, as demonstrated by the K_{rel} for Fr3. The association rate of the BS15-RNA complex ($1x10^4$ $M^{-1}s^{-1}$) is much slower than that expected for a diffusion controlled process (typically around $1x10^8$ $M^{-1}s^{-1}$), as was observed for the ribosomal S8 protein (Mougel *et al.*, 1986). This is in contrast to the interaction between various tRNAs and their cognate synthetases, for which rates of association are diffusion controlled (Krauss *et al.*, 1973; Pingoud *et al.*, 1973; Pingoud *et al.*, 1975).

The slow rate of association is likely due to a requirement for a collision in the correct orientation for specific binding, where not all protein-RNA collisions are productive. The association rate is dependent on protein concentration, suggesting that the rate limiting step of binding is a bimolecular association. However, this data does not rule out the possibility that a rate limiting unimolecular conformational change must occur in the RNA prior to formation of a stable complex. In fact, for many enzyme-substrate and protein-protein interactions, observed association rate constants which are significantly slowere than diffusion controlled can be attributed to a two-step process appearing as a single step (Fersht, 1985). Under the conditions commonly utilized to measure the association rate contant, the reaction between an enzyme and a substrate in which substrate binding is preceded by a unimolecular conformational change in the enzyme, the kinetic equations adopt the form of a simple bimolecular association (Fersht, 1985). This problem can only be resolved by measuring the kinetics at high substrate concentrations (in this case, high RNA concentrations, which is impractical using the gel-shift technique). However, there is evidence that just such a unimolecular conformational change is occurring in the RNA at some point during S15 binding. The enhanced mobility of Fr1

RNA bound to S15 suggests that a conformational change in the RNA does accompany protein binding. Fr1, Fr4, and Fr5 all display an enhancement of mobility in a native polyacrylamide gel upon binding to BS15 (Figure 3.4), despite the fact that the protein is highly basic and adds 10 kDa of mass to the complex, either of which would be expected to decrease an RNA's mobility in a gel. This suggests that S15 binds to its site at the junction of three helices and organizes the helices in a particular orientation with respect to one another. This gel mobility enhancement has also been observed with the RNA binding proteins HIV rev and alfalfa mosaic virus coat protein, and attributed to a conformational change in the RNA upon protein binding (Kjems *et al.*, 1992; Baer *et al.*, 1994). All other minimal sites which bind BS15 with high affinity show a decrease in electrophoretic mobility upon S15 binding, which is most likely due to the shortened helix 21 in those RNAs.

The conformational change in the RNA upon S15 binding suggests a role for this protein within ribosomal assembly. *In vitro* reconstitution experiments of the *E. coli* 30S subunit indicate that the binding of the S6/S18 heterodimer to the central domain is strongly facilitated by the binding of S15. The assembly pathway suggests that S15 is the key primary binding protein for the incorporation of S6, S18, S11, and S21 into the central domain of the ribosome. The facilitation of S6/S18 binding could be accomplished through two means. First, the S6/S18 binding site is directly adjacent to S15 as indicated by Fe^{2+}-EDTA studies, suggesting that cooperativity could be mediated by protein-protein interactions. There is also the possibility that S15 could facilitate the assembly of this domain through an RNA conformational change by organizing helices 20, 21, and 22 with respect to one another, creating the binding site in the 16S rRNA for S6:S18. It seems likely that the conformational change induced in the three-way RNA junction by S15 is a major contributor to its facilitation of the assembly of the central domain of the 30S ribosomal subunit. This is also consistent with electron microscopy data demonstrating a

distinct condensation of rRNA structure upon binding of many of the primary binding proteins, including S15 (Mandiyan *et al.*, 1989).

Chapter 4: Analysis of the S15 binding site in rRNA

In the preceding chapter, it was demonstrated that the *Bacillus stearothermophilus* ribosomal protein S15 interacts with a 61 nucleotide minimal rRNA domain with high affinity. This minimal RNA site contains an internal loop with two phylogenetically conserved A•G base pairs and a junction between three duplex helices. Two of the helices (helices 20 and 21) are necessary primarily to maintain the integrity of the three-way junction, and that most of the sequence-specific recognition occurs in the third helix (helix 22). In this chapter, the elements of specificity of the BS15-16S rRNA interaction were determined using site specific mutagenesis, chemical modification interference, and iodine footprinting, in order to verify the validity of the minimized protein-RNA complex as a good model for structural studies.

Site specific mutations in Fr1 RNA

To test the specificity of the BS15-Fr1 RNA interaction, a representative set of mutations were introduced into Fr1 RNA, targeting the internal loop and junction regions within helix 22, taking into consideration the prokaryotic 16S rRNA phylogeny. Seven mutants of Fr1 RNA were constructed (Figure 4.1) and their relative affinities for BS15 as compared to wild-type Fr1 were measured using a competition assay (Table 4.1). Four of the five nucleotides in the three-way junction (U652, G654, G752, and A753 in Fr1 RNA) are highly conserved in prokaryotic, archaeal, and chloroplast phylogeny (Gutell, 1994). In the few species which do exhibit changes at these positions in 16S rRNA, there is no covariation with any of the other conserved nucleotides that might suggest base pairing. In

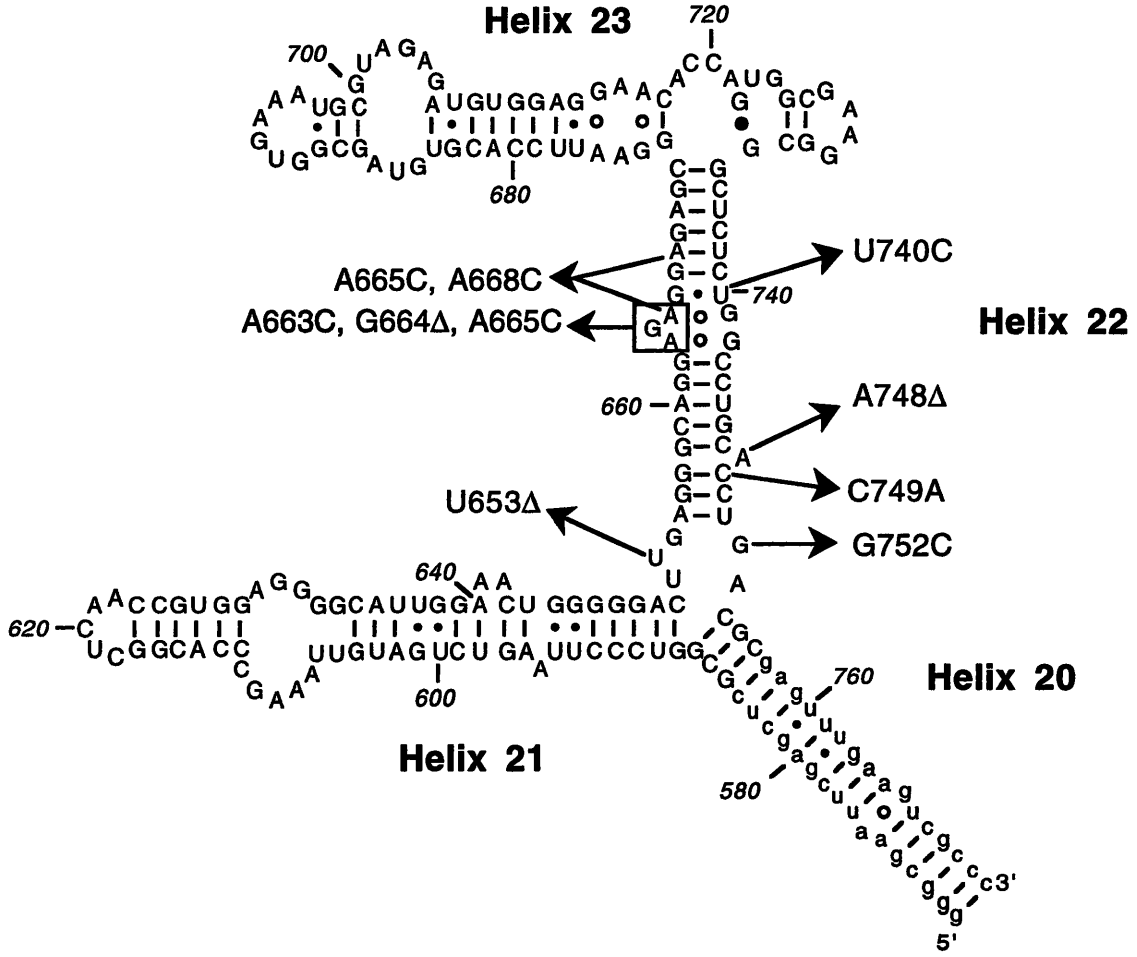


Figure 4.1: Secondary structure of Fr1 RNA, which is derived from the central domain of *Bacillus sterothermophilus* 16S rRNA, and the mutations introduced into this RNA for this study. The numbering of the RNA and the helices is consistent with the conventional numbering of *Escherichia coli* 16S rRNA. Nucleotides which are not wild-type to *B. stearothermophilus* 16S rRNA are denoted in lower case letters.

Table 4.1: Relative binding affinities of Fr1 RNA mutants.

Fr1 RNA mutant	K_{rel}*
Fr1(G752C)	72
Fr1(U653Δ)	0.70
Fr1(A748Δ)	30
Fr1(C749A)	1.1
Fr1(A663C, G664Δ, A665C)	0.62
Fr1(A665C, A668C)	0.91
Fr1(U740C)	5.7

*$K_{rel}=(K_{d,mutant}/K_{d,Fr1})$

this element, G752 was mutated to a cytidine, creating a potential Watson-Crick base pair between G654-C752. The affinity of BS15 for this mutant was reduced 72-fold, which is almost half of the maximal loss in affinity observed in the series of deletion mutants. Substitution of an adenosine at position 752 in the context of a minimal site also showed a strong decrease in BS15 affinity for the RNA (data not shown). These results demonstrate that this guanosine residue is a critical component of the BS15-rRNA interaction. In contrast to the other four nucleotides in the junction, U653 was extremely variable across phylogeny; all four nucleotides were observed with high frequency at this position. A point deletion, U653Δ, was made to test the possible role of this nucleotide as a "spacer" required for BS15 binding. Deletion of U653 did not lead to any loss in affinity of BS15 for this RNA, and therefore, this nucleotide must be important for another role in the 30S subunit. Furthermore, like Fr1 RNA, titration of Fr1(U653Δ) RNA with BS15 yielded a higher mobility complex in an 8% native polyacrylamide gel (see Figure 5.1), indicating that the conformational change occurring in the three-way junction can be accommodated by the four phylogenetically conserved nucleotides.

Another conserved feature near the junction is a single bulged nucleotide in helix 22. Although always present, this nucleotide is variable in both position and identity throughout prokaryotic phylogeny. All four nucleotides are found at this position and the position of the bulged nucleotide varies between three and five base pairs away from the bottom of helix 22. The bulged A748 was deleted to determine its role in S15-RNA recognition. The resulting mutant, Fr1(A748Δ) exhibited a thirty-fold reduction in BS15 affinity. However, the introduction of the C749A mutation immediately downstream of this important nucleotide, which should disrupt a Watson-Crick base pair, had little effect upon protein binding. This demonstrates the importance of the bulged nucleotide in helix 22 for the S15-16S rRNA interaction.

A number of studies have suggested that S15-rRNA recognition occurs at a phylogenetically conserved internal loop containing two A•G base pairs and a single

bulged guanosine (Müller *et al.*, 1979; Mougel *et al.*, 1988; Svensson *et al.*, 1988; Powers & Noller, 1995). As a drastic alteration of this site, a mutation was made such that the internal loop was replaced with two Watson-Crick C-G base pairs in Fr1(A663C, G664Δ, A665C). Surprisingly, this mutant, along with Fr1(A665C, A668C) which should also significantly disrupt the internal loop, binds to BS15 with wild type affinity. Instead, BS15 appears to recognize the conserved G667-U740 pair immediately above the internal loop. The conversion of this wobble pair into a standard Watson-Crick G-C base pair leads to a moderate drop in affinity, consistent with the reduction in binding affinity observed in Fr6, which contained a deletion of this region. These results implicate the G-U base pair, and not the internal loop, as a primary determinant of the BS15-rRNA interaction around the internal loop.

Identification of important nucleotides in Fr1 for BS15 binding

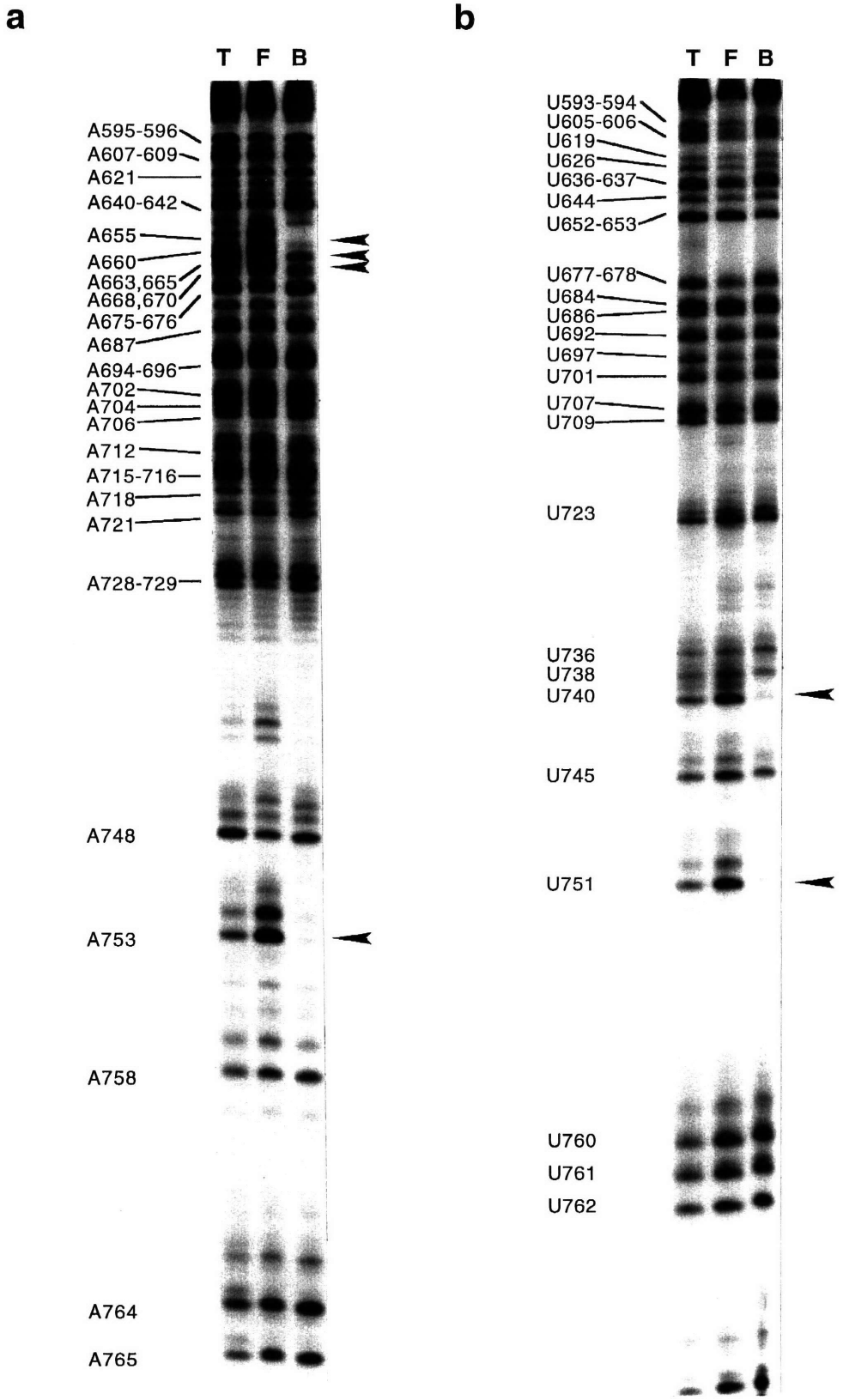
The chemical modification interference assay was used to identify nucleotides important for the BS15-Fr1 RNA interaction. 3' end-labeled Fr1 RNA was probed with either diethyl pyrocarbonate (DEPC), which carboxyethylates the N7 positions of adenosine (and guanosine to a lesser extent), or hydrazine under aqueous conditions, which depyrimidinates uracil bases. The RNA was modified under denaturing conditions, such that all potential sites were modified at a low level with approximately equal frequency. The modified RNA was incubated with a concentration of BS15 that yielded approximately 50% binding, where discrimination between various sites of modification in the RNA should be most sensitive. Positions where modification interferes with protein binding were visualized by the comparison of free and bound RNA on a denaturing sequencing gel following cleavage of the RNA at the sites of modification. The interference of protein binding by a modification at a given position could be due to

disruption of either a specific protein-RNA contact or an RNA structure necessary for protein binding.

Interferences of protein binding due to modification of moieties in Fr1 RNA by DEPC or aqueous hydrazine are clearly seen in Figure 4.2. Carboxyethylation of A655 and A753, which are located near the three-way junction, are the strongest modification interferences observed. A753 is one of the four nucleotides in the junction that is virtually invariant across prokaryotic phylogeny, and A655 is part of an A-U base pair adjacent to the single-stranded nucleotides at the junction that also displays strong phylogenetic conservation. Likewise, U751, which is base paired to A655, is a strong interference, reinforcing the importance of the integrity of this base pair for protein recognition. On the other hand, A748, which mutational analysis indicated was important for protein binding, shows no effect on protein binding upon modification. Weaker modification interferences were observed in A663 and A668. This contrasts the two mutants, Fr1(A663C, G664Δ, A665C) and Fr1(A665C, A668C), which indicates these adenosines have a negligible impact upon protein binding. Thus, modifications at these positions are likely to perturb the RNA structure rather than disrupt direct protein-RNA contacts. U740 demonstrates a strong modification interference, consistent with the observed drop in binding affinity of the Fr1(U740C) mutant. Furthermore, neither set of modifications show any interferences outside of the protein binding site as defined by deletion analysis, lending further validity to the minimal RNA binding site as a good model substrate for the BS15-16S rRNA interaction.

Modification of the highly conserved U652 in the single stranded junction element does not appear to interfere with protein binding. However, the adjacent uridine could potentially substitute for the modified uridine. To test this, the modification interference

Figure 4.2 (next page): Modification interference assays of 3' end-labeled Fr1 RNA using DEPC (a) and aqueous hydrazine (b) as the RNA modifying agents. Lanes are labeled as T for total modified RNA that was not subjected to protein binding, F (free) for RNA that did not bind BS15, and B (bound) for the RNA that bound BS15. Positions whose modification interfered with protein binding are denoted with arrowheads.


a
T F B
A595-596
A607-609
A621
A640-642
A655
A660
A663,665
A668,670
A675-676
A687
A694-696
A702
A704
A706
A712
A715-716
A718
A721
A728-729
A748
A753
A758
A764
A765
b
T F B
U593-594
U605-606
U619
U626
U636-637
U644
U652-653
U677-678
U684
U686
U692
U697
U701
U707
U709
U723
U736
U738
U740
U745
U751
U760
U761
U762

assay was repeated using Fr1(U653Δ) RNA. This RNA has the same affinity as wild type RNA, however, it contains only one uridine in the single stranded region near the three-way junction. The pattern of DEPC and aqueous hydrazine modification interferences for this RNA was almost identical to that of Fr1 RNA (Figure 4.3) The only difference is that depyrimidination of U652 now becomes a strong interference to BS15 binding in the context of Fr1(U653Δ) RNA, revealing the importance of this uridine at the three-way junction.

Interference analysis of BS15-Fr7 RNA interaction

Modification interference provides an extremely sensitive method for demonstrating whether the specificity of the BS15-rRNA interaction has been preserved in the course of creating a minimal binding substrate. Fr7 RNA binds to BS15 with wild type affinity, despite deletion of most of helix 21 and all of helix 23. However, the binding measurement alone does not directly demonstrate that the mode of specificity for protein binding to this RNA is identical. The patterns of RNA modification interference of protein binding provides detailed information on the specific nucleotides which are intimately involved in the interaction, and thus is a good qualitative measurement of the binding specificity. Modification interference experiments performed on Fr7 RNA using DEPC (Figure 4.4a) and aqueous hydrazine (data not shown) demonstrate very similar patterns of interferences between Fr1 and Fr7. This convincingly demonstrates that helix 23 and the distal portion of helix 21 provide no elements of specificity necessary for the S15-rRNA interaction. The only notable difference between Fr1 and Fr7 is that the importance of U740 of the G-U base pair is somewhat diminished relative to Fr1 RNA.

Ethylnitrosourea (ENU), which ethylates the phosphate oxygens in the backbone, was also used to probe the BS15-Fr7 interaction. ENU modification interference needed to

Figure 4.3 (next page): Modification interference assays of 3' end-labeled Fr1(U653Δ) RNA using DEPC (a) and aqueous hydrazine (b) as the RNA modifying agents. Lanes are labeled as T for total modified RNA which was not subjected to protein binding, F (free) for RNA that did not bind BS15, and B (bound) for the RNA that bound BS15. Positions whose modification interfered with protein binding are denoted with arrowheads.

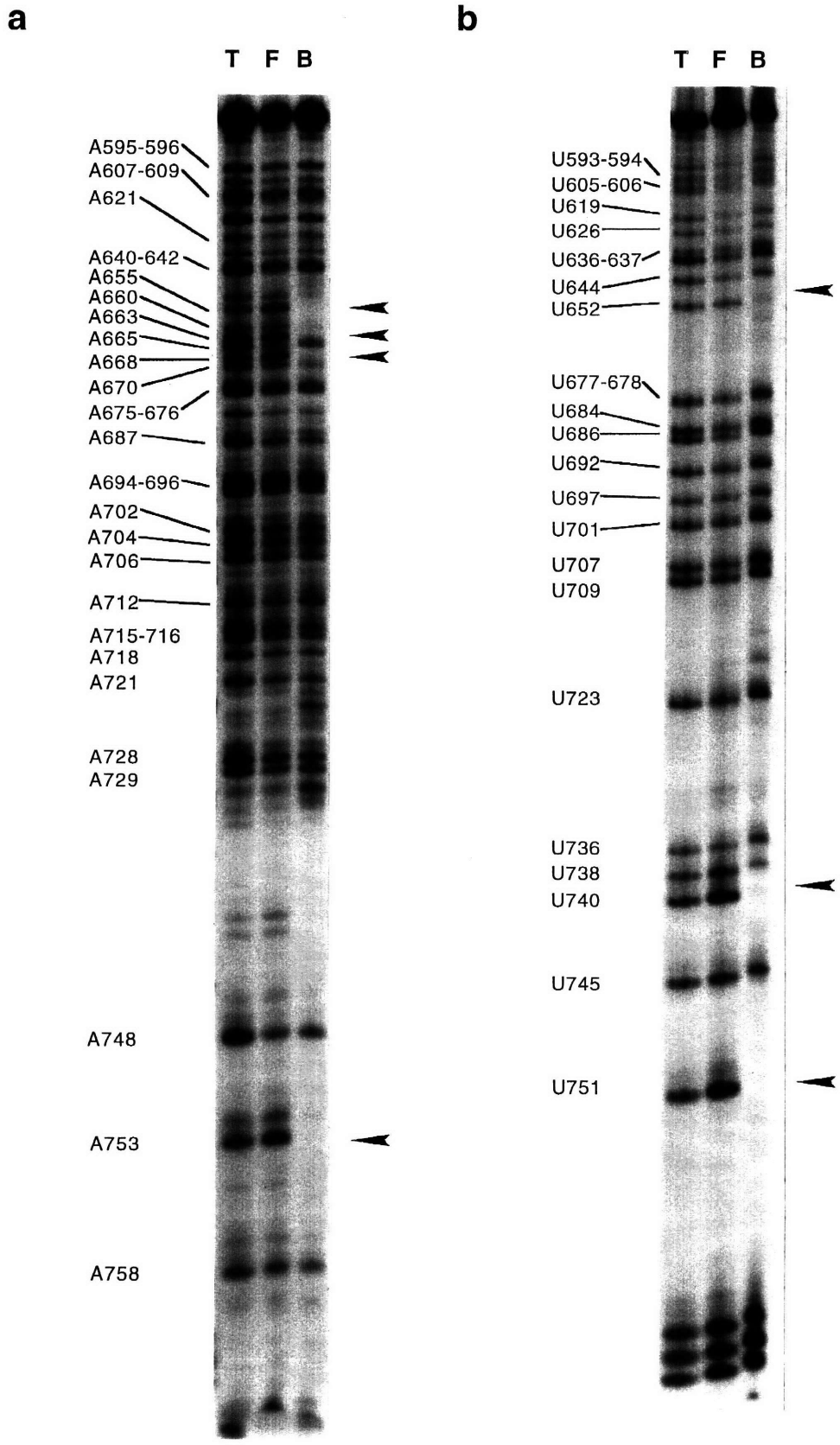
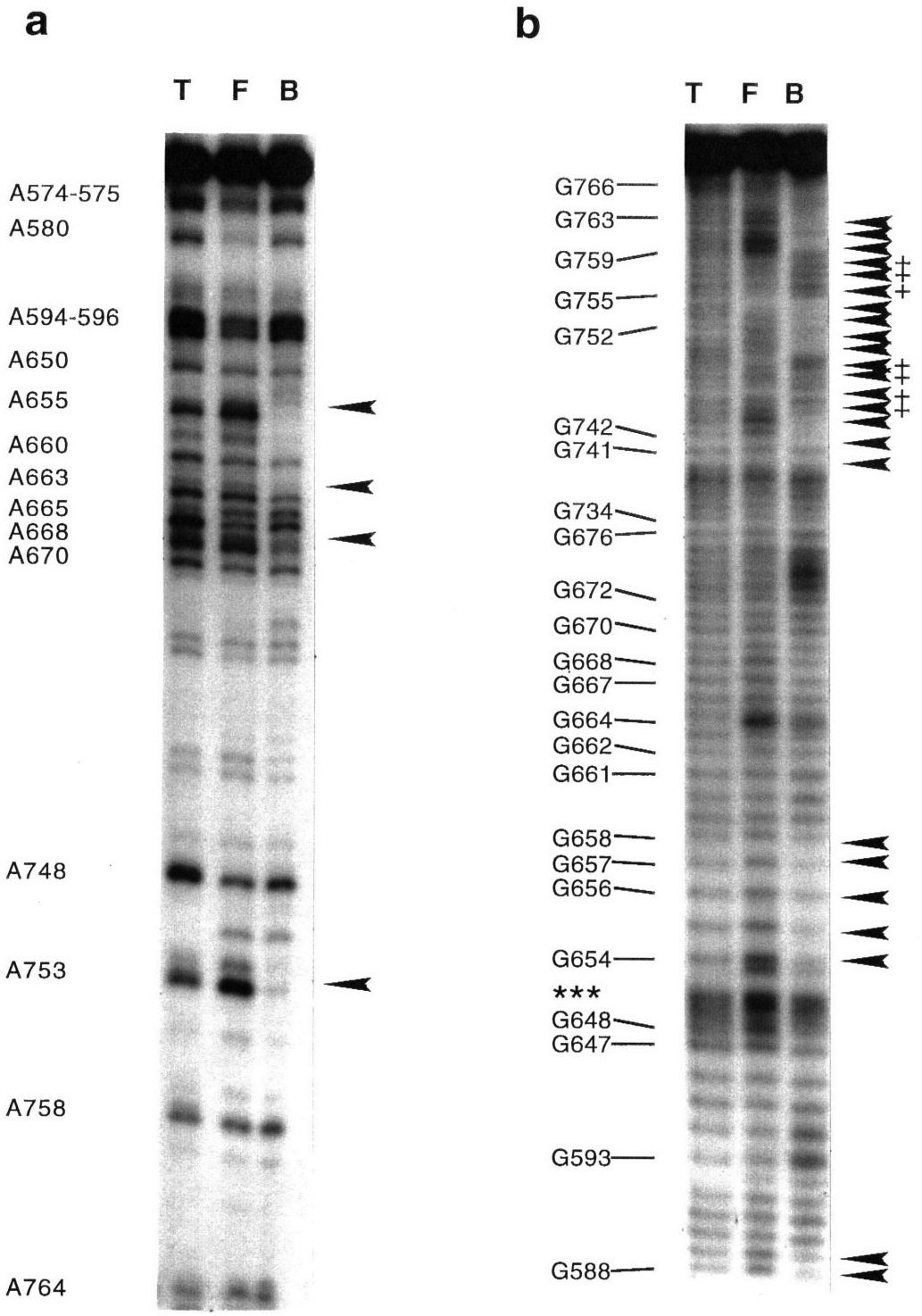

a
T F B
A595-596
A607-609
A621
A640-642
A655
A660
A663
A665
A668
A670
A675-676
A687
A694-696
A702
A704
A706
A712
A715-716
A718
A721
A728
A729
A748
A753
A758
b
T F B
U593-594
U605-606
U619
U626
U636-637
U644
U652
U677-678
U684
U686
U692
U697
U701
U707
U709
U723
U736
U738
U740
U745
U751

Figure 4.4 (next page): Modification interference assays of BS15 binding to Fr7 RNA using DEPC modified 3' end-labeled RNA (a) and ENU modified 5' end-labeled RNA (b). The bands in the ENU lanes were identified by alkaline hydrolysis and RNase T_1 ladders run in adjacent lanes (not shown). Interferences are denoted with an arrowhead, enhancements are denoted with an arrowhead and a plus sign, and band compressions and other gel artifacts are denoted with asterisks.


a
T F B
A574-575
A580
A594-596
A650
A655
A660
A663
A665
A668
A670
A748
A753
A758
A764
b
T F B
G766
G763
G759
G755
G752
G742
G741
G734
G676
G672
G670
G668
G667
G664
G662
G661
G658
G657
G656
G654
***
G648
G647
G593
G588

be performed on both 5' and 3' end-labeled Fr7 RNA, since significant compression artifacts were observed near the tetraloops (starred nucleotides in Figure 4.4b) due to folding back of the RNA in these regions. 3' end-labeled RNA was able to resolve the compression observed around G649-U753 in 5' end-labeled Fr7 RNA (data not shown), revealing a strong modification interference at the 5' phosphate of G654. In addition, both 5' and 3' labeled RNA consistently showed a set of very strong modification interferences at the 5' phosphates of C743 and C744.

Beginning with the phosphate interferences at C743, a distinct pattern of modifications and enhancements were observed towards the 3' end of Fr7 RNA. In this portion of the RNA, comparison of the band intensities between the free and bound RNA lanes shows an alternating pattern of interferences and enhancements (Figure 4.4, b). The switch between enhancements and interferences in the 3' region occurs every five to six nucleotides, which is approximately one half turn of an A-form RNA helix. The periodicity of this pattern clearly indicates that BS15 is contacting one side of the RNA helix along an extended region between C743 and A764 in helices 20 and 22.

Detailed analysis of the S15-rRNA determinants

Modification interference analysis was performed in detail on Fr15 RNA, one of the smallest RNAs which could still bind BS15 with wild type affinity. The strength of each interference was determined as the intensity difference between a given band in the free RNA compared to that in the bound RNA, and the results are summarized in Table 4.2. DEPC and hydrazine (under conditions which depyrimidinate at uridines) shows a pattern of interferences virtually identical to those seen in RNA Fr1 and Fr7, demonstrating that the specificity of the BS15-rRNA interaction is maintained in Fr15 (Figure 4.5). Modification of Fr15 using hydrazine under conditions which predominantly depyrimidinate at cytidine residues reveals interferences above the three-way junction at C749 and C750 and at C739

Table 4.2: Summary of BS15 dependent modification interferences and footprints of Fr15 RNA.

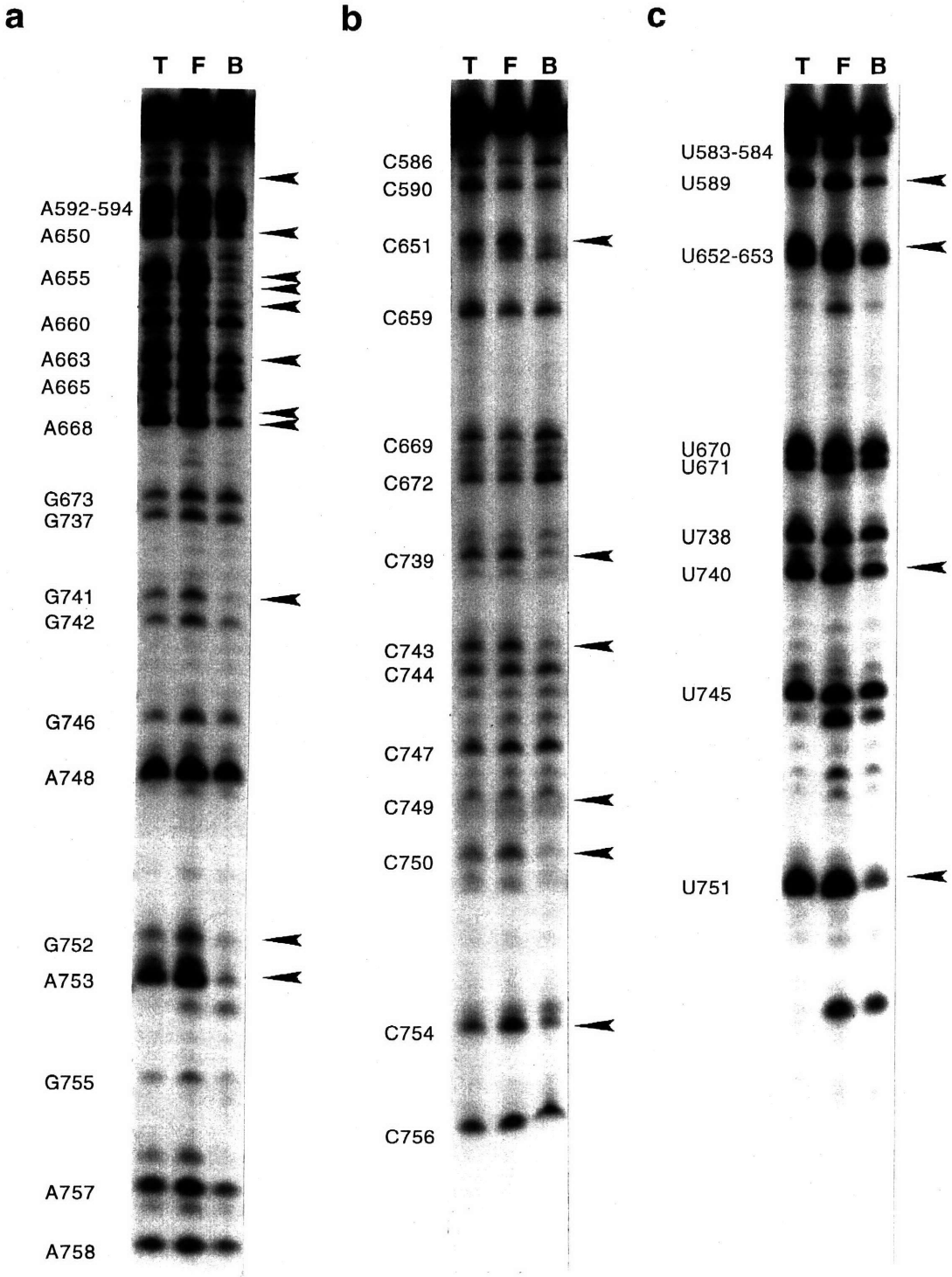
Nucleotide	DEPC	Hydrazine	ENU	CMCT	DMS[a]	Iodine
G580	n.d.		n.d.	n.d.		
G581	n.d.		n.d.	n.d.		
G582	n.d.		n.d.	n.d.		
U583		o	o	o		
U584		o	o	o		
G585	o		o	o		
C586		o	o			
G587	+		o	+		
G588	++		++	+++		
U589		++	++	+++		+
C590		+	o			
G591	n.d.		o	o		
A592	o		o		o	
A593	*		o		o	
A594	o		o		o	
G649	n.d.		o	+		
A650	++		o		o	
C651		+++	o			+
U652		+++	+	+++		+++
U653		+	++	o		++
G654	++		+++	+++		+
A655	+++		++		+++	+++
G656	n.d.		++	+++		++
G657	++		++	+++		++
G658	+		o	+++		++
C659		+	o			+
A660	+		o		o	
G661	+		o	++		
G662	+		o	++		
A663	+++		o		o	
G664	o		o	o		+
A665	+		o		o	+++
G666	+		o	o		
G667	++		o	+		
A668	+++		o		o	
C669		o	o			
U670		*	o	o		
U671		o	o	o		
C672		o	o		o	
G673	o		o	o		
G737	o		o	o		
U738		o	o	o		

C739		+++	o			
U740		+++	++	+++		*
G741	+++		++	o		
G742	++		++	o		+++
C743		+++	+++			+++
C744		+	+++			+++
U745		o	*	+		
G746	o		*	o		
C747		o	*			
A748	o		*		o	
C749		+	*			
C750		+++	*			++
U751		+++	++	+++		++
G752	+++		++	o		+++
A753	+++		++		o	+++
C754		+++	++			+
G755	+		+	o		+
C756		o	*			
A757	n.d.		*		o	
A758	n.d.		n.d.		o	
C759		n.d.	n.d.			
C760		n.d.	n.d.			
C761		n.d.	n.d.			

In the above table the following one letter code was used: (+++) strong interference, (++) medium interference, (+) weak interference, (*) enhancement, (o) no effect, and (n.d.) for positions that could not be monitored.

[a]With the exception of C672, DMS did not significantly modify C(N3) positions.

Figure 4.5 (next page): Modification interference assays of BS15 binding to 3' end-labeled Fr15 RNA using DEPC (a), hydrazine under high salt and anhydrous conditions (b), and aqueous hydrazine (c) as RNA modification agents. Interferences are denoted with an arrowhead.


a
T F B
A592-594
A650
A655
A660
A663
A665
A668
G673
G737
G741
G742
G746
A748
G752
A753
G755
A757
A758
b
T F B
C586
C590
C651
C659
C669
C672
C739
C743
C744
C747
C749
C750
C754
C756
c
T F B
U583-584
U589
U652-653
U670
U671
U738
U740
U745
U751

and C743 around the purine-rich internal loop. Two other strong cytidine modification interferences occurs at C651 and C754 in helices 21 and 20 respectively.

Along with these two cytidines, there are a number of nucleotides in helices 20 and 21 of Fr15 whose modification interfere with BS15 binding that are not apparent in RNAs Fr1 or Fr7. Moderate and strong interferences are observed in most of the nucleotides of the two base pairs adjacent to the three-way junction of helices 20 and 21. This supports the observation that the stability of these helices is necessary for high affinity binding of BS15. The base pair adjacent to the three-way junction in both helices was conserved throughout the minimization process described previously, and thus could potentially provide contacts to BS15. However, the G587-C754 pair of helix 20 showed a different pattern of modification interference from that of the G588-C651 pair of helix 21. While both nucleotides of the base pair adjacent to the three way junction in helix 21 showed significant interferences to protein binding upon modification, only C754 of the base pair adjacent to the junction in helix 20 showed strong modification interferences. Furthermore, mutation of G587 to an adenosine in the context of a minimal RNA binding site could still bind BS15 with wild type affinity. Phylogeny also demonstrates an asymmetry to the conservation pattern of these two nucleotides; while C754 is highly conserved, a uracil is often found at the 787 position to generate a C•U mismatch. Thus, there is a strong possibility that in the nucleotides at the 587 and 754 position do not pair in the context of the S15 bound rRNA, and that the conserved C754 is playing some other role in RNA structure or protein recognition. The identity of the second base pair in each helix has been shown to be unimportant for the high affinity BS15-RNA interaction. Thus, these interferences do not represent direct protein contacts, but rather, disruptions in RNA structure which inhibit protein binding.

The pattern of interferences observed with ENU using 5' and 3' end-labeled RNA is similar to that observed for Fr7 RNA, including the pattern of enhancements and interferences seen towards the 3' end of the RNA (Figure 4.6). In these experiments

modifications corresponding to the 5' phosphates of U653 through G657 disappear in the free RNA in both 5' and 3' end-labeled RNA, making this region difficult to interpret. However, the strong interference observed at the 5' phosphate of G654 is consistent with Fr7 data as well as weaker interferences in the region between U652 and G658.

The role of the imino positions of the bases in this protein-RNA interaction was probed by modification of the A(N1) and C(N3) positions with DMS and the G(N1) and U(N3) positions with CMCT. These modifications were visualized by reverse transcription of the modified RNA since modifications on the Watson-Crick face inhibit reverse transcriptase. These experiments were performed using Fr15/*Eae*I RNA, which is Fr15 RNA with a 17 nucleotide single stranded tail at its 3' end. Prior to these experiments, it was determined that Fr15/*Eae*I RNA binds BS15 with wild type affinity as determined by direct titration (data not shown). Therefore, the introduction of this single-stranded region does not alter the RNA structure or its ability to bind BS15.

The only significant modification interference around the internal loop occurs at U740 of the G-U base pair (Figure 4.7). Interestingly, there is no interference at its complement, G666, despite the bulky modification placed at this position. The two base pairs above the G-U pair that show strong interferences when modified by DEPC and hydrazine display no interference from modification of the imino position. However, G661 and G662 which display weak DEPC modification interferences, show strong modification interferences due to CMCT. Another set of strong interferences occurs in helix 22 in the three base pairs adjacent the three-way junction. The A655-U751 base pair is particularly intolerant to modification of either the adenosine or the uridine, indicating that Watson-Crick pairing of these nucleotides is essential for the BS15-RNA interaction. Modification interferences at base pairs proximal to the three-way junction in helix 20 and 21 further demonstrate the need for intact base pairs in these helices to maintain the integrity of the junction.

Figure 4.6 (next page): Modification interference assays of BS15 binding to ENU modified Fr15 RNA using 5' end-labeled (a) and 3' end-labeled RNA (b). The individual lanes in each series are (1) T_1 digested Fr15 RNA, (2) alkaline hydrolysis ladder, (3) total ENU modified RNA, (4) free Fr15 RNA, and (5) bound Fr15 RNA. Modification interferences are denoted with arrowhead and modification enhancements are denoted with arrowheads with plus signs.

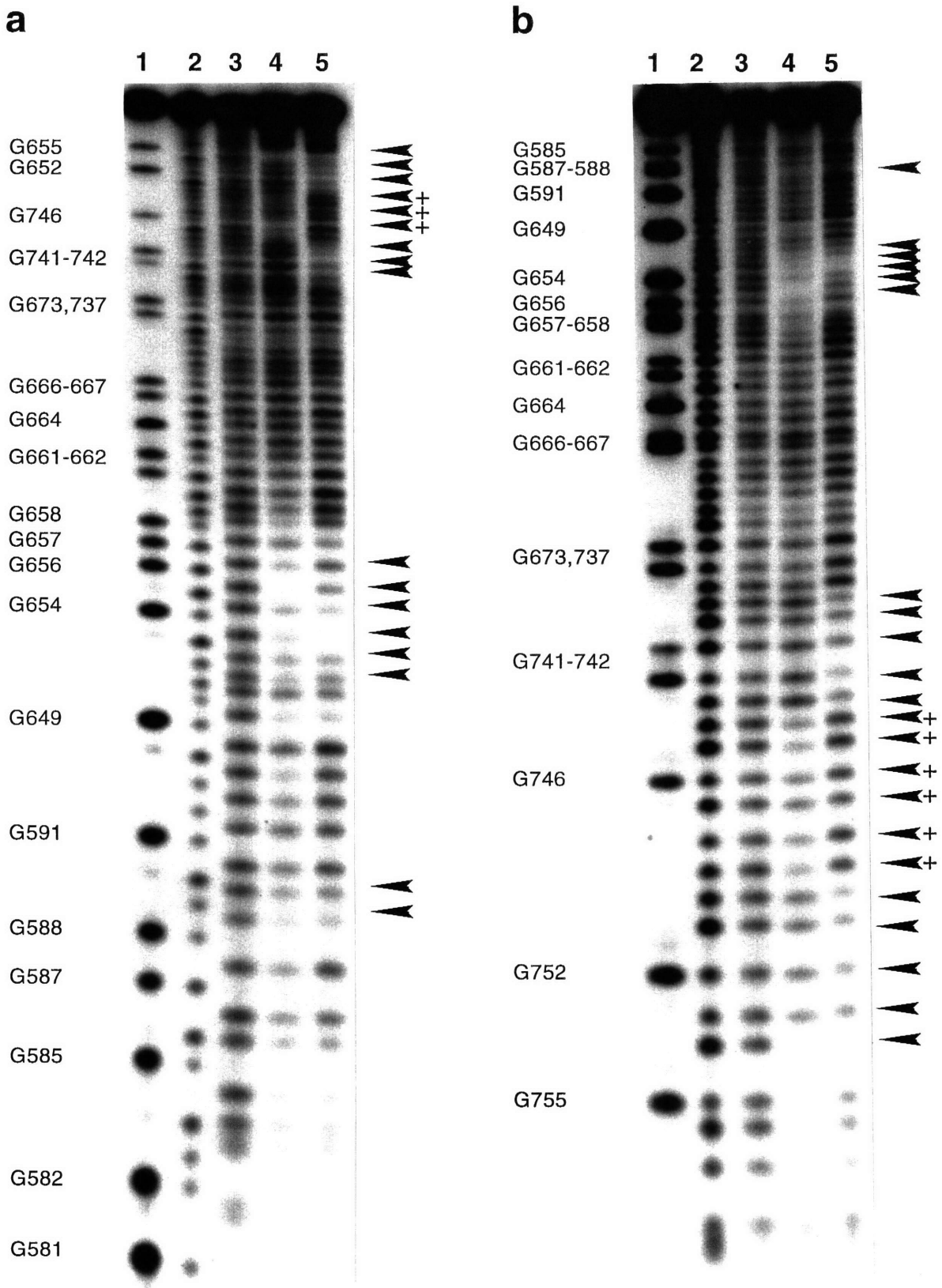
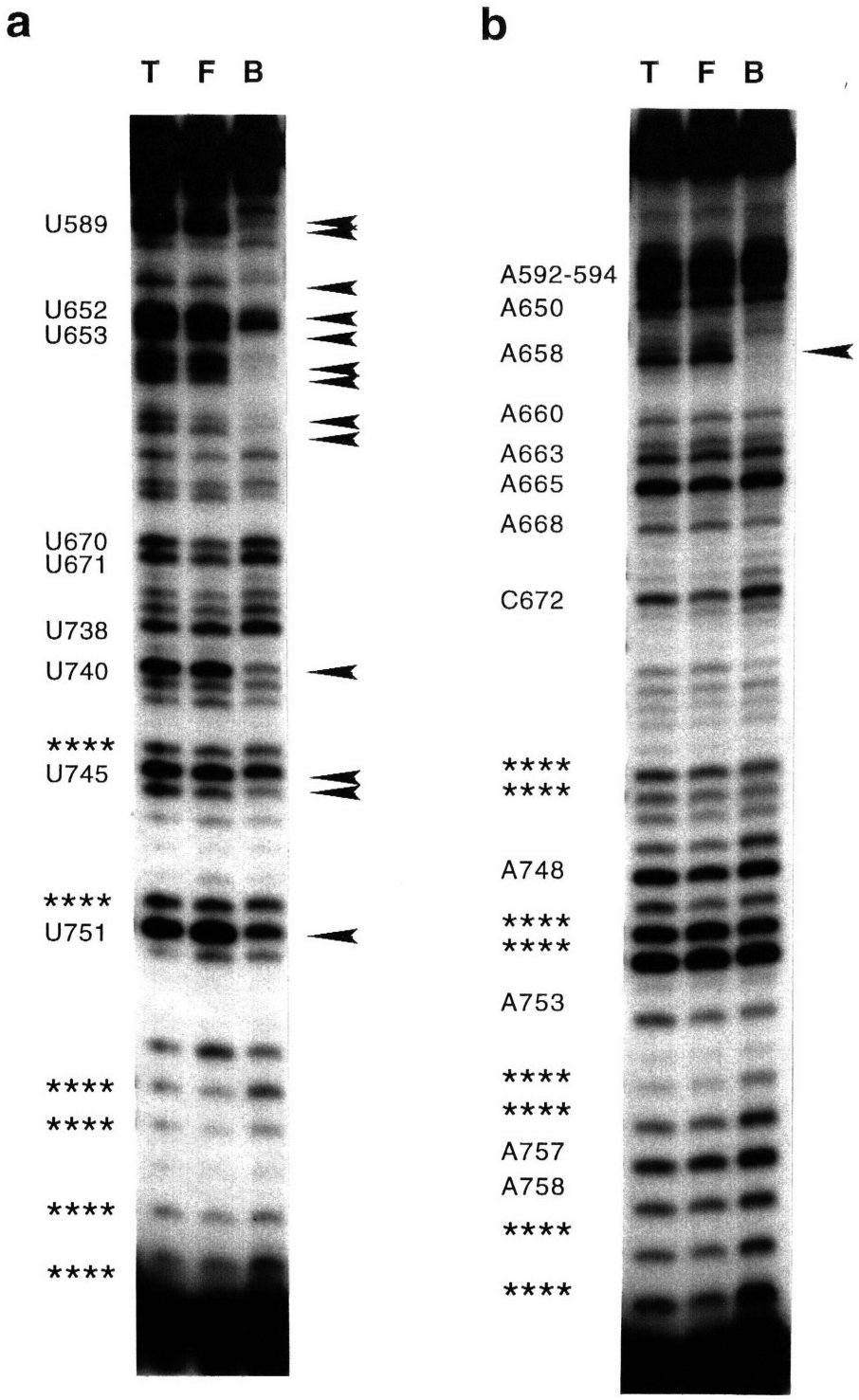

a
1 2 3 4 5
G655
G652
G746
G741-742
G673,737
G666-667
G664
G661-662
G658
G657
G656
G654
G649
G591
G588
G587
G585
G582
G581
+++
b
1 2 3 4 5
G585
G587-588
G591
G649
G654
G656
G657-658
G661-662
G664
G666-667
G673,737
G741-742
G746
G752
G755
+
+
+
+
+
+

Figure 4.7 (next page): Modification interference assays of BS15 binding to Fr15/*Eae*I RNA using CMCT (a) and DMS (b) as RNA modification agents. Lanes are marked as (T) for total RNA, (F) free RNA, and (B) bound RNA. Nonspecific reverse transcriptase stops are denoted with asterisks.


a
T F B
U589
U652
U653
U670
U671
U738
U740
****
U745
****
U751
****
****
****
****
b
T F B
A592-594
A650
A658
A660
A663
A665
A668
C672
****
****
A748
****
****
A753
****
****
A757
A758
****
****

In the single-stranded region near the three-way junction, G752 and A753, which are both extremely sensitive to modification of their N7 positions, show no interference due to modification at the N1 position. In contrast, U652 and G654 both display strong interferences from modification of their Watson-Crick face. The importance of the Hoogsteen face of G752 and A753 and the Watson-Crick face of U652 and G654 leads us to suggest that these four nucleotides form two non-canonical base pairs in which the Watson-Crick face of U652 and G654 form pairing hydrogen bonds with the Hoogsteen face of A753 and G752, respectively, as shown in Figure 4.8. These four nucleotides are virtually invariant in prokaryotic 16S rRNA, and as a result there is no covariation available to provide phylogenetic evidence for the presence of these base pairs. However, the conservation of these four nucleotides at the three-way junction, mutagenesis, and the observed modification interferences suggest that they may play a critical role in the RNA structure.

Footprinting of the BS15-Fr15 interaction

Iodine cleavage of phosphorothioate-containing Fr15 RNA transcripts was used as a means to directly footprint the BS15-Fr15 interaction. An earlier study of Powers and Noller demonstrated that the *E. coli* ribosomal protein S15 footprinted to 16S rRNA in the lower part of the 660/750 stem, near the three-way helical junction, and to the purine-rich internal loop using the Fe^{2+}-EDTA reaction to probe the RNA (Powers & Noller, 1995). In order to directly demonstrate that the *E. coli* and *B. stearothermophilus* variants of S15 were recognizing similar RNA elements, we probed the specific BS15-Fr15 complex using a similar footprinting technique. Footprinting of protein-RNA interactions using iodine to cleave phosphorothioate containing RNAs has been used as an equivalent method to

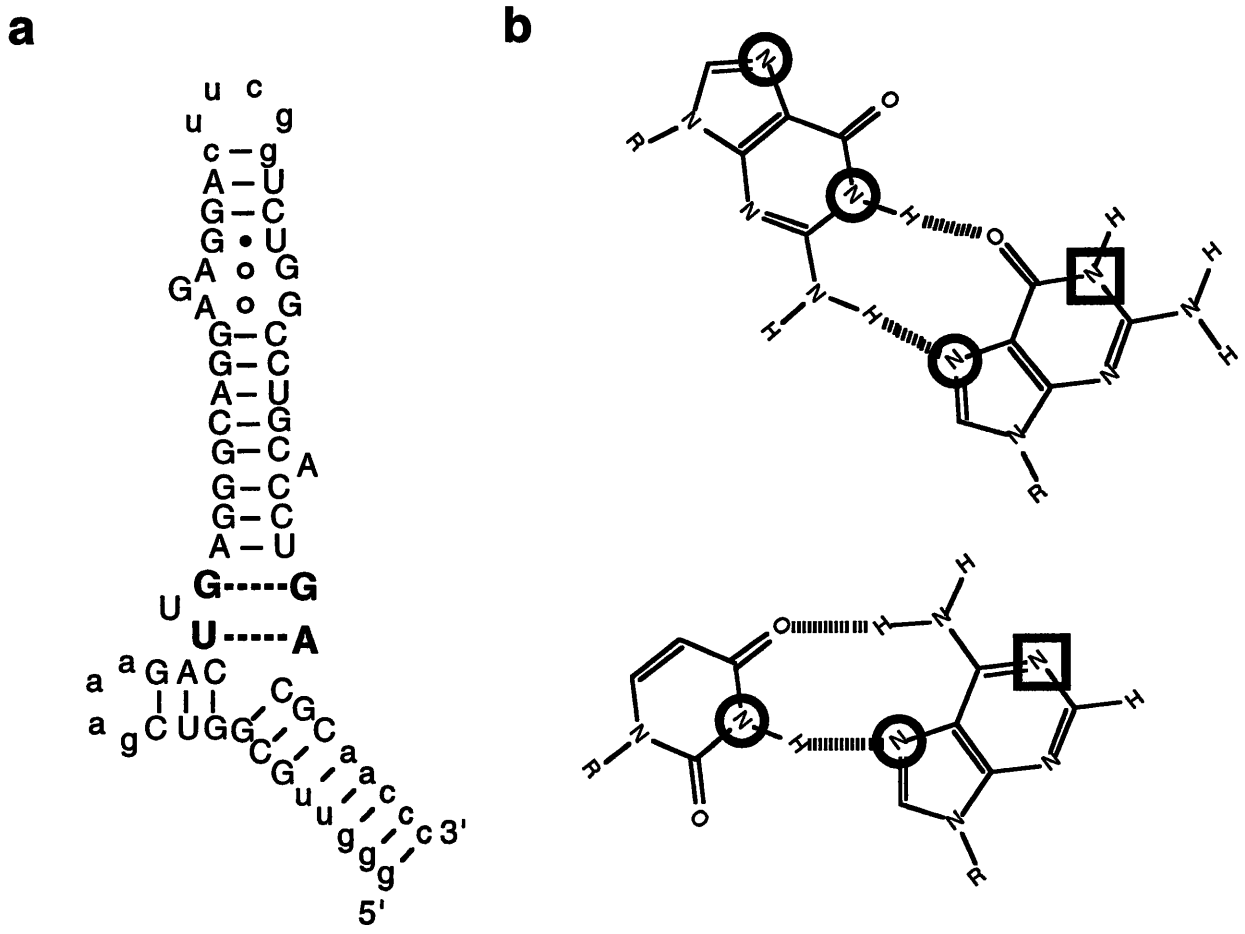
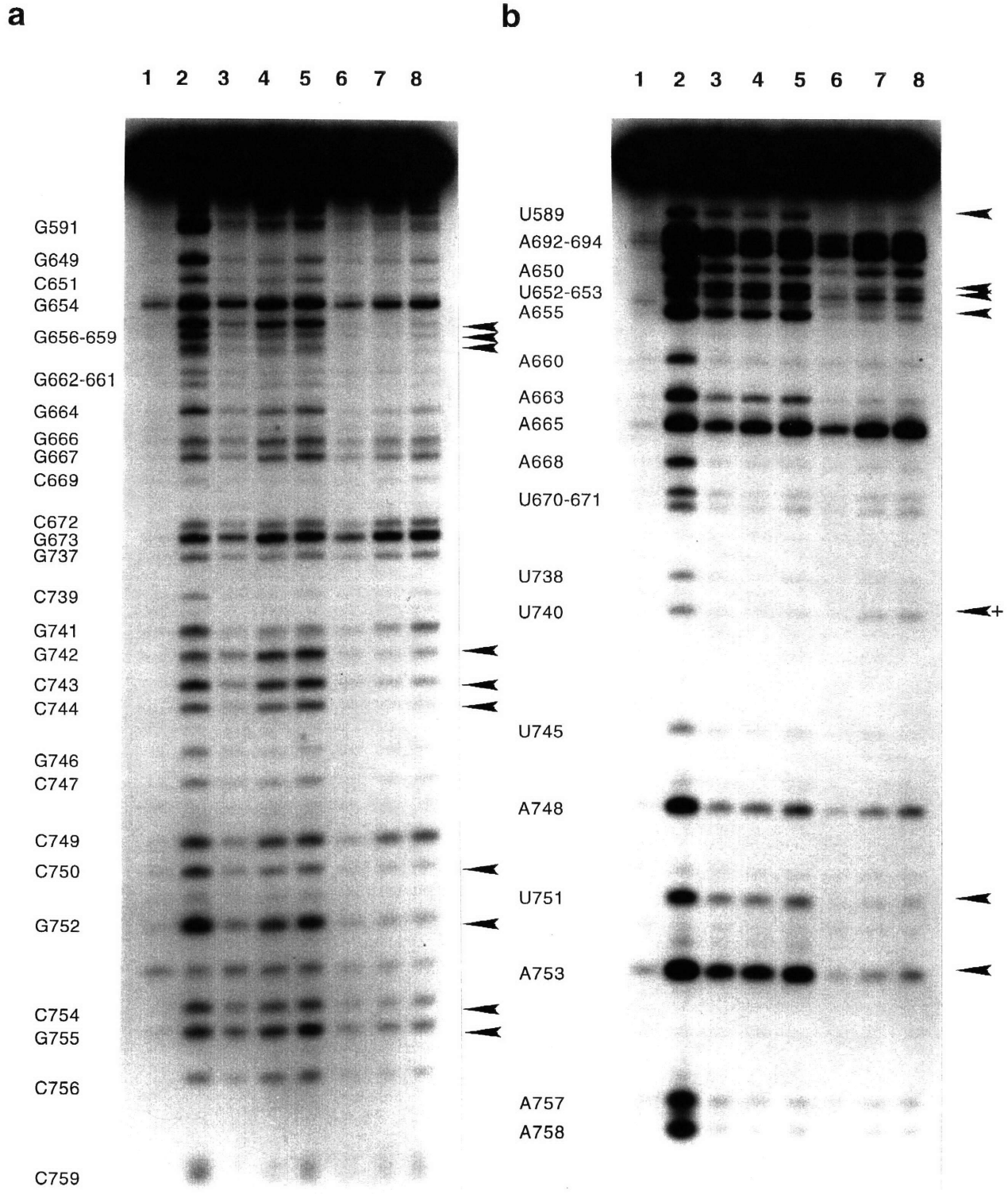


Figure 4.8: Proposed base pairs near the three-way junction as suggested from modification interference assays. (a) Location of the proposed G654-G752 and U652-A753 base pairs in Fr15 RNA. (b) Pairing scheme of these base pairs. The circles denote positions whose modification interferes with protein binding, and squares denote positions whose modification does not interfere with protein binding.

localize regions of contact between the protein and RNA (Schatz *et al.*, 1991; Baron *et al.*, 1993).

Footprinting was carried out on native Fr15 RNA and the BS15-Fr15 complex with varying concentrations of I_2 (Figure 4.9). Probing was performed with various iodine concentrations (100 μM, 500 μM and 1 mM) to visualize sites of weak and strong protection (Schatz *et al.*, 1991). In the absence of protein, most of the phosphates demonstrate approximately the same degree of cleavage under native and denaturing conditions, although the degree of cleavage is quite variable between nucleotides. Upon addition of 200 nM BS15, a number of phosphates were strongly protected from cleavage by iodine, as summarized in Table 4.2. The pattern of interferences strongly correlates with the reported Fe^{2+}-EDTA footprint observed for the *E. coli* S15-16S rRNA interaction and the ethylation modification interference analysis performed using BS15. The only significant difference between the protection patterns of the two proteins is near the internal loop containing the A•G base pairs. *E. coli* S15 (ES15) protects the upper base pair of the internal loop as well as two base pairs above it. The BS15 protection in this region centers on the lower base pair of the internal loop and the base pair below it. This observation is consistent with ENU modification interference experiments in which ethylation of the phosphates 5' to C743 and C744 interferes strongly with protein binding. However, both ES15 and BS15 demonstrate strong modification interferences with the three base pairs above the internal loop. This convincingly demonstrates that these two proteins are recognizing 16S rRNA in a highly homologous fashion and most likely recognizing the same RNA elements.

Figure 4.9 (next page): Iodine cleavage protection assays using 3' end-labeled Fr15 RNA doped with 5% phosphorothioate guanosine and cytosine (a) or adenosine and uracil (b). Individual lanes are (1) phosphorothioate-labeled RNA not subjected to iodine cleavage, (2) cleavage under denaturing conditions, (3) free RNA cleaved with 100 μM iodine, (4) free RNA cleaved with 500 μM iodine, (5) free RNA cleaved with 1 mM iodine, (6) bound RNA cleaved with 100 μM iodine, (7) bound RNA cleaved with 500 μM iodine, and (8) bound RNA cleaved with 1 mM iodine.


a
1 2 3 4 5 6 7 8
G591
G649
C651
G654
G656-659
G662-661
G664
G666
G667
C669
C672
G673
G737
C739
G741
G742
C743
C744
G746
C747
C749
C750
G752
C754
G755
C756
C759
b
1 2 3 4 5 6 7 8
U589
A692-694
A650
U652-653
A655
A660
A663
A665
A668
U670-671
U738
U740
+
U745
A748
U751
A753
A757
A758

Discussion

Interpretation of the determinants of the BS15-rRNA interaction

The mutagenesis, chemical modification interference, and footprinting data provide a detailed picture of the region of the 16S rRNA that interacts with BS15. These data are compared with the phylogenetic conservation of each nucleotide in the S15 recognition site of 16S rRNA for eubacteria and related 16S rRNAs in Figure 4.10. The nucleotides that are important for the BS15-16S rRNA interaction show a striking correspondence to the highly conserved nucleotides. The correlation between the biochemical data and the 16S rRNA phylogeny suggests that the features observed for the *B. stearothermophilus* S15-16S rRNA interaction may be general for prokaryotes, archae, and chloroplasts. The S15 protein also shows a high degree of homology throughout eubacterial and archaeal phylogeny, further supporting the idea that the essence of the S15-rRNA interaction is conserved across these kingdoms.

The 16S rRNA binding site of S15 is composed of two distinct sites of recognition. The first site of interaction is centered around an internal loop containing two phylogenetically conserved A•G base pairs. Previous work has suggested that the primary elements of S15 recognition in this region are the purine-purine base pairs of the internal loop (Müller *et al.*, 1979; Mougel *et al.*, 1988; Svensson *et al.*, 1988), however, our data demonstrate that the primary sequence determinant of the S15-rRNA interaction in this region is the conserved G-U pair immediately above the internal loop. S15 is certainly making other contacts to this region, since mutation of the G-U pair to a G-C pair does not account for all of the binding energy lost upon deletion of this region. However, the mutagenesis and modification interference data suggest contacts to the phosphates of C743

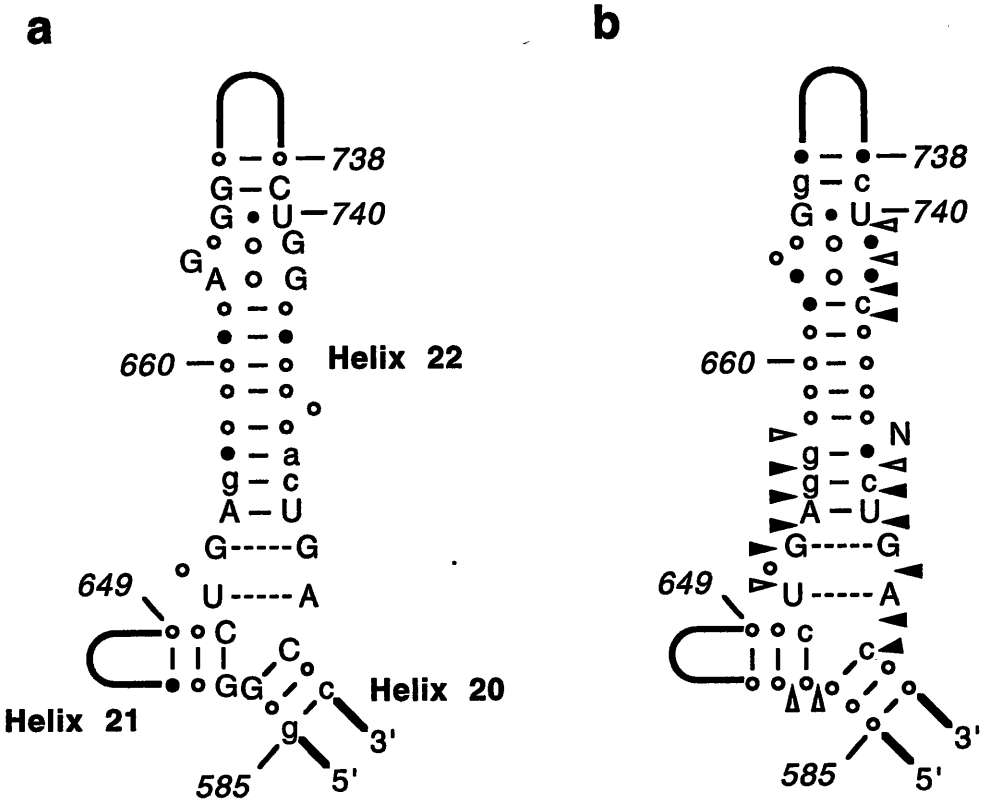


Figure 4.10: (a) Phylogenetic conservation of nucleotides in Fr15 RNA. Nucleotides in capital letters are greater than 95% conserved across eubacteria and chloroplasts, lower case letters represent nucleotides that are 90-95% conserved, filled dots represent nucleotides that are 80-90% conserved, and open dots represent nucleotides that exist relative to *E. coli* 95% of the time (*i.e.* the existence of a nucleotide is conserved at this position, but the identity of the nucleotide is not conserved) (adapted from the 16S ribosomal RNA database, Gutell, 1994). (b) Summary of the mutational and modification interference data from Fr15 RNA. Nucleotides in capital letters are essential for BS15 recognition as determined by both mutagenesis and modification interference, nucleotides which are in lower case show some strong modification interferences, filled dots represent nucleotides that show medium/weak modification interferences, and open dots represent nucleotides that are present in all of the RNAs studied, but whose identities are not important for the BS15-rRNA interaction.

and C744, but do not support other potential contacts in this region.

The fact that S15 does not appear to recognize the most prominent conserved secondary structural motif in helix 22 of 16S rRNA is somewhat surprising. However, various biochemical studies have implicated other proteins that interact in this region of the rRNA. The binding of the S6/S18 heterodimer to the assembling 30S subunit is directly potentiated by S15, and results in the protection of the bulged G664 from attack by kethoxal, and portions of the internal loop from Fe^{2+}-EDTA cleavage (Svensson *et al.*, 1988; Powers & Noller, 1995). Thus, the cooperative binding of S6/S18 to the 16S rRNA potentially could be due to the influence of S15 on the structure of the rRNA in the region of the internal loop, as well as possible protein-protein contacts. Also, mRNA has been shown to be in very close proximity to the internal loop by crosslinking to the adenosine base of the upper A•G pair (Jutta *et al.*, 1994). This reinforces the idea that the internal loop is not highly sterically occluded by many direct contacts with S15.

Another significant determinant of the BS15-rRNA interaction occurs in helix 22 between the internal loop and the three-way junction. Across eubacterial phylogeny, the presence of a bulged nucleotide in this helix is highly conserved, but any nucleotide can be accommodated, and the position of this bulged nucleotide can be either three or four base pairs away from the three-way junction. Deletion of this bulged nucleotide has a very strong effect upon the binding activity of BS15, but chemical modification of the N7 position or N1 imino does not affect binding. Therefore, the presence of a bulged nucleotide plays a structural role in the RNA binding site, but does not appear to be directly recognized by the protein.

The region that displays the strongest determinants of the BS15-rRNA interaction lies in helix 22 proximal to the three-way junction. Modification interference and mutational analysis clearly implicate six nucleotides as critical for the RNA-protein recognition. In all of the RNAs tested, any modification to the A655-U751 base pair immediately above the three-way junction strongly diminishes protein binding. In addition,

four nucleotides (U652, G654, G752, and A753) that are classically drawn as single-stranded nucleotides in the 16S rRNA and are extremely well conserved throughout eubacterial phylogeny, are implicated in the RNA-protein interaction. Strong modification interferences on the Watson-Crick face of U652 and G654 and on the Hoogsteen face of G752 and A753 lead us to propose that they form two unusual base pairs. The extreme conservation of these putative base pairs suggests that these base pairs might uniquely define an RNA backbone geometry which is recognized and stabilized by the protein and allows the three helices to be oriented in a specific conformation.

To illustrate how BS15 is interacting with helix 22 of 16S rRNA, the data generated from the mutagenesis, modification interference, and iodine footprinting experiments were projected onto a three-dimensional representation of this helix (Figure 4.11). In this model, helix 22 of Fr15 RNA was built using standard A-form geometry to define the parts of the helix which contain standard Watson-Crick base pairs. The unpaired nucleotides U653, G664, and A748 were bulged out of the helix, and the UUCG tetraloop was built according to the NMR model of this element (Cheong *et al.*, 1990). The non-canonical A•G pairs were represented in the side-by-side conformation that has been found in many studies of model RNA oligonucleotides containing these pairs by NMR (SantaLucia & Turner, 1993; Katahira *et al.*, 1994), although we have no direct evidence that these nucleotides are paired in this fashion. The two proposed non-canonical base pairs at the three-way junction have also been included in this model.

This model presents a clear picture of how S15 is recognizing this helix of 16S rRNA. The critical base pairs A655-U751 and G666-U740 are separated by one complete turn of the A-form RNA helix, clearly indicating that S15 recognizes two distinct sites in the rRNA, both situated on the same face of the helix. Both of these sites are flanked by phosphates that are protected by S15 in iodine footprinting experiments and whose ethylation interferes with S15 binding. These phosphates are situated on the minor groove

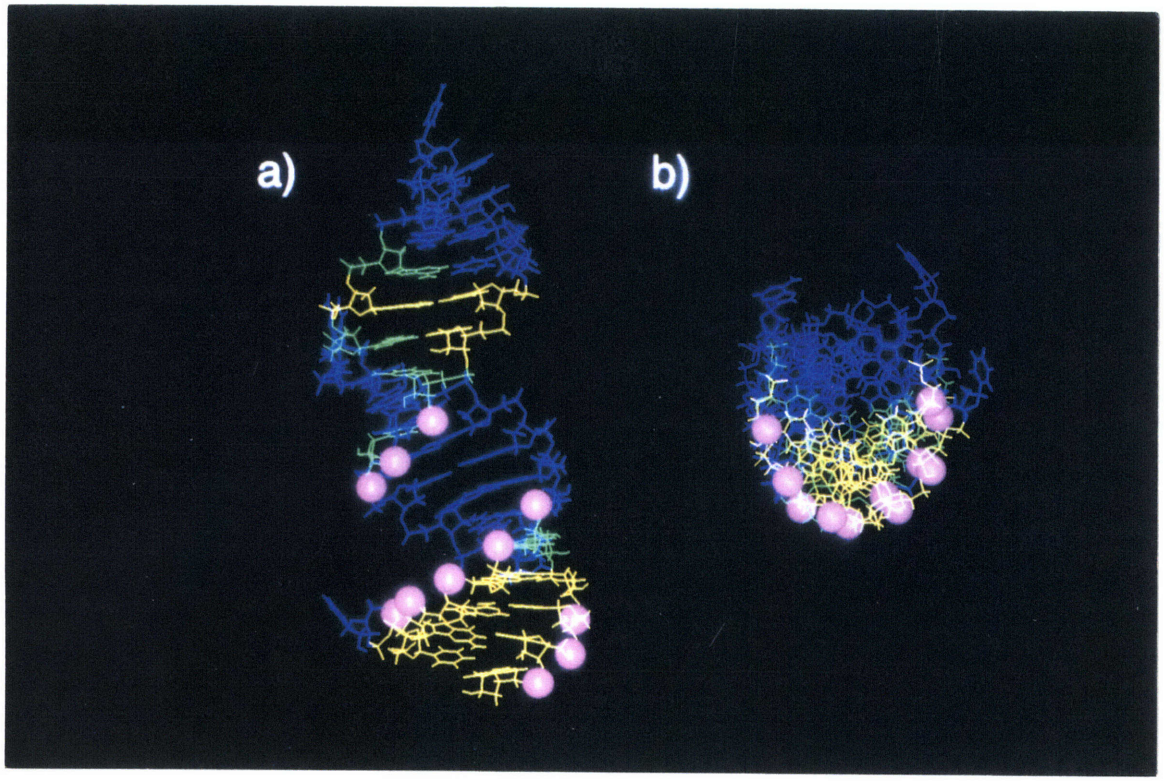


Figure 4.11: The three-dimensional model of helix 22 of Fr15, with the mutation, modification interference, and iodine footprinting data superimposed upon it. The two views show (a) the minor groove of the two regions that directly interact with BS15 and (b) the model looking down the helical axis. The phosphates whose ethylation interferes with protein binding and that are protected from iodine cleavage with a phosphorothioate substituted at that position are shown as purple spheres. Nucleotides of weak or moderate importance as determined by mutagenesis and modification interference are shown in green, and the most important nucleotides are shown in yellow.

side of both of the base pairs, along one face of the helix. Therefore, S15 appears to be recognizing the rRNA in the minor groove of helix 22 near the three-way junction and crossing over the major groove in the middle of this segment of the helix to recognize further elements in the minor groove around the G-U pair. The important bulged adenosine, which is not directly contacted by the protein, lies on the opposite face of the helix, away from the sites of direct protein recognition. This nucleotide is probably situated in the major groove on the opposite face, influencing the groove widths and RNA geometry in the protein binding site near the junction.

Many of these elements of the BS15-16S rRNA interaction are found in other RNA-protein interactions. A well documented example of recognition of a G-U pair via interactions with the minor groove is the $tRNA^{Ala}$-AlaRS interaction. It has been demonstrated that a single G-U base pair in the acceptor stem of $tRNA^{Ala}$ is the major determinant for aminoacylation of that tRNA by *E. coli* alanine tRNA synthetase (Hou & Schimmel, 1988; McClain & Foss, 1988). Studies of the ability of AlaRS to aminoacylate model duplexes containing specific chemical substitutions clearly implicates the guanosine exocyclic amino group of the wobble G-U pair as critical for aminoacylation (Musier-Forsyth *et al.*, 1991; Musier-Forsyth & Schimmel, 1992). Non-canonical nucleotide base pairing is also a recurring motif used by proteins in recognition of RNA. A well studied example of unusual base pairing in rRNA is the loop E motif, found originally in eukaryotic 5S rRNA, which provides the binding site for the proteins TFIIIA (Romaniuk, 1989; Allison *et al.*, 1991) and L25 (Huber & Wool, 1984; Zhang & Moore, 1989). NMR studies of oligonucleotides containing the loop E motif reveal that a side-by-side anti-anti G•A pair and a reverse-Hoogsteen anti-anti A•U pair is formed by highly phylogenetically conserved nucleotides (Szewczak *et al.*, 1993; Wimberly *et al.*, 1993). Unusual base pairing has also been implicated at RNA helical junctions that are recognized by ribosomal proteins, such as L11. This protein has been shown to bind to a phylogenetically

conserved junction between three helices in the large subunit 23S rRNA, and has been shown to require a specific folding of this junction, part of which is an unusual A•U base pair (Ryan & Draper, 1989; Ryan *et al.*, 1991). Tertiary base pairs at the junction of multiple helices in domain III of 23S rRNA have also been shown to be critical for recognition by ribosomal protein L25 (Kooi *et al.*, 1993).

The conformation of the rRNA three-way junction

The binding of S15 to the minor groove of helix 22 of 16S rRNA induces a significant conformational change, in which the spatial orientation of helices 20, 21, and 22 changes with respect to one another. The mobility enhancement in native polyacrylamide gels of select RNAs upon binding BS15 and the pattern of phosphate ethylation modification interferences in Fr7 RNA both strongly suggest that helices 20 and 22 are nearly parallel to one another. The strong interferences observed at C743 and C744, and U760-A764 are quite distant from each other; approximately 60 Å apart by tracing along the two helices. The pattern of interferences and enhancements along helix 20 and 22 suggest they may be brought into a conformation such that they are almost parallel to one another, as diagrammed in Figure 4.12. Since each of these sets of interferences are approximately ten base pairs away from the three-way junction, if helices 20 and 22 are oriented parallel to one another at the junction, it would place these two sets of interferences very close to one another in a way that BS15 could easily contact these distant sites.

Models of the 16S rRNA in the ribosomal 30S subunit have been built independently by several groups using data from numerous biochemical and biophysical studies of the 30S subunit structure (Brimacombe *et al.*, 1988; Stern *et al.*, 1988c; Malhotra & Harvey, 1994). The 30S models clearly place helices 20 and 22 in a conformation such that they are nearly running parallel to one another and in close proximity. The extreme distal loops of helices 23 and 24 are easily crosslinked together and are both implicated in

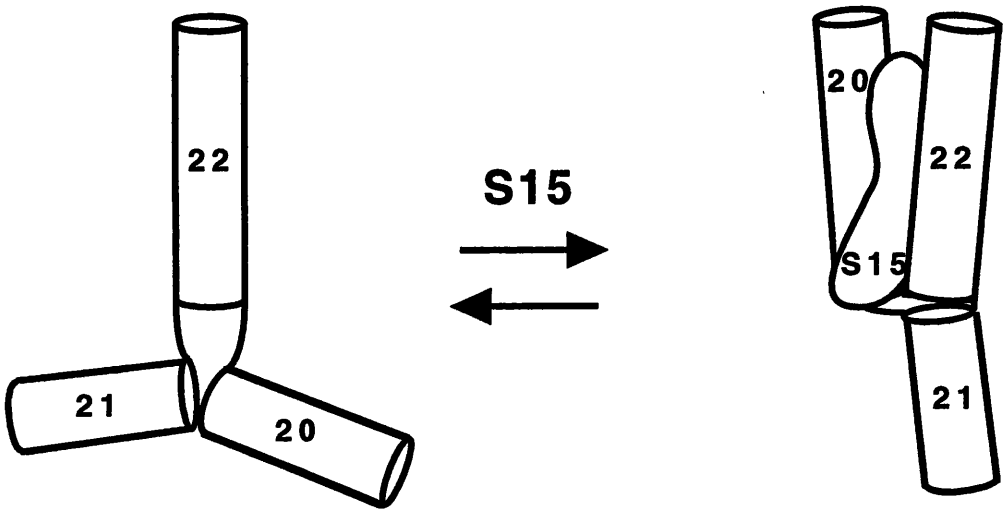


Figure 4.12: Cartoon of the conformational change induced about the RNA three-way junction upon S15 binding, as suggested by native polyacrylamide gel electrophoresis and by ethylation modification interference.

the P-site of the 30S subunit (Stern *et al.*, 1988c; Malhotra & Harvey, 1994). Helices 23 and 24 are believed to stack upon helices 22 and 20, respectively, and thus, the orientation of the these helices at the three-way junction may be ultimately responsible for bringing these distant RNA elements into close proximity. The 30S models generally place helix 21 at a large obtuse angle with respect to helices 20 and 22, but do not suggest that helices 20 and 21 coaxially stack, as might be expected. Thus, the more global picture of the 30S structure is supportive of the detailed biochemical picture of the S15 binding region.

Relationship between the rRNA and mRNA binding site of S15

It has been demonstrated that S15 recognizes a structure in the 5' untranslated region of its own mRNA that is in equilibrium between a conformation containing a series of three stem-loop structures and a pseudoknot conformation (Portier *et al.*, 1990b; Portier *et al.*, 1990c). S15 binds to the pseudoknot conformation of this mRNA and promotes the assembly of a stalled pre-initiation complex, thus preventing its own translation (Phillipe *et al.*, 1993). Footprinting and mutagenesis of this mRNA has led to a detailed picture of S15 binding to the pseudoknot element (Benard *et al.*, 1994; Phillipe *et al.*, 1995). A comparison of the Fe^{2+}-EDTA footprint of *E. coli* S15 bound to CFP5517 mRNA (a mutant of the wild-type RNA in which the mutation has stabilized the pseudoknot conformation) and the iodine footprint of *B. stearothermophilus* S15 bound to helix 22 of 16S rRNA is shown in Figure 4.13.

There are a number of features of these two sites that are similar. In both the mRNA and rRNA, S15 appears to directly interact with a G-U base pair (shown boxed in Figure 4.13). In the mRNA, this base pair appears to be the major sequence specific determinant, whereas in the rRNA site, it is not as important as determinants near the junction. The S15 footprint of these two sites also shows strong similarities. Both sites show two distinct regions with which S15 interacts, each spaced almost one turn of the

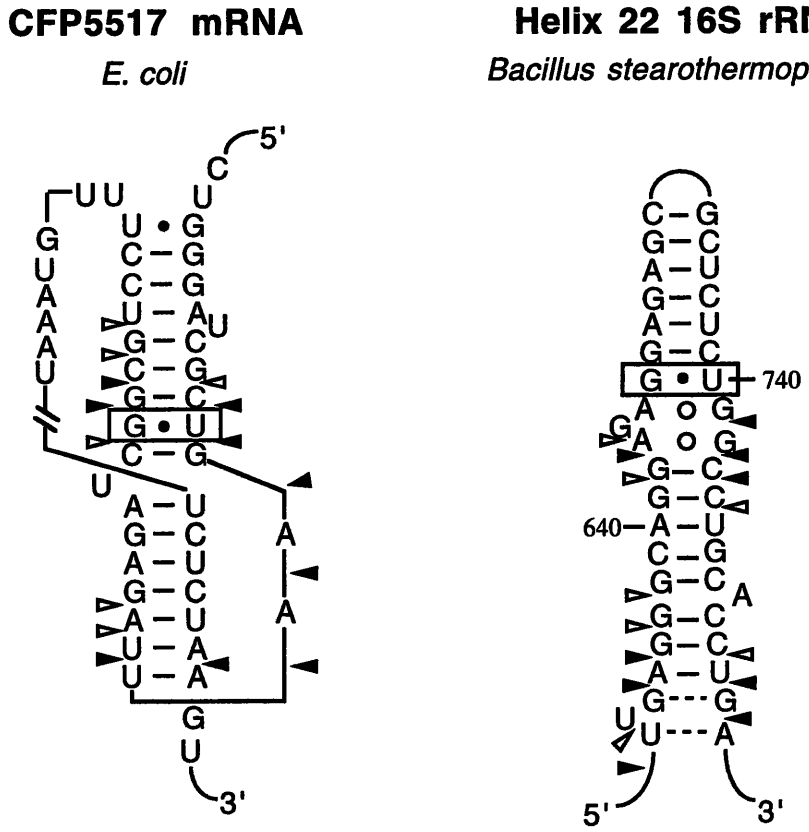


Figure 4.13: Comparison of the ENU footprint of ES15 bound to CFP5517 RNA (a mutant of the S15 translational operator in which the pseudoknot conformation has been stabilized) (adapted from Phillipe *et al.*, 1995) and the iodine footprint of BS15 bound to helix 22 of *B. stearothermophilus* rRNA. Sites of strong protection are denoted as filled arrowheads and sites of weak protection are denoted as open arrowheads. The G-U basepair important for recognition of the mRNA and rRNA by S15 is boxed.

RNA helix away from one another. In each site there is a distinct shift in the pattern of protections in the 3' direction on opposite sides of the helix, indicative of a protein interacting with the minor groove (Dervan, 1986; Oakley & Dervan, 1990; Lee *et al.*, 1991). Therefore, in both the mRNA and the rRNA, S15 likely recognizes elements in the minor groove along an extended face of one helix, including a common G-U base pair.

In a proposed model of the pseudoknot structure bound by S15, the loop adenosines of the pseudoknot, which are strongly protected from chemical modification by S15, were placed in the major groove on the same face of the helix to which S15 binds (Phillipe *et al.*, 1995). This differs from the model of the rRNA site, in which the bulged adenosine is placed opposite the S15 binding face. However, these single stranded nucleotides in both cases may be critical for inducing a backbone conformation that can be recognized by S15. In both the mRNA and rRNA site, data suggest that the RNA sugar-phosphate backbone provides determinants for S15 recognition. Therefore, we expect that minor groove recognition of the mRNA pseudoknot and helix 22 of 16S rRNA to be facilitated by a similar RNA backbone geometry created by two different RNA motifs.

Note in proof: After the studies presented in Chapters 3 and 4 were published (Batey & Williamson, 1996a; Batey & Williamson, 1996b), another study was published detailing the interaction between *Thermus thermophilus* S15 and 16S rRNA, using many of the same techniques and methods (Serganov *et al.*, 1996). This study determined that TS15 binds to a nearly identical minimal site in 16S rRNA on the minor groove face of helix 22, but their interpretation of the potential base pairing scheme in the phylogenetically conserved three-way junction element, the orientation of the helices about the junction motif, and the role of magnesium was slightly different (discussed in Chapter 5).

Chapter 5: Folding of the RNA three-way junction

Using a combination of chemical probing and mutagenesis, it has been demonstrated that a critical feature of the S15 recognition site in 16S rRNA is a three-way junction motif. Upon binding S15, this junction undergoes a significant conformational change which reorients the three flanking helices. Other three-way junction motifs, most notably the hammerhead ribozyme, have been shown to be capable of a similar, specific folding event, facilitated by the presence of polyvalent cations. Whether the folding of the rRNA junction requires S15 or can be similarly facilitated by other cations is the focus of this chapter.

Nucleic acid junctions

The structure of large RNAs is determined in most cases by the structure of smaller RNA elements that serve as architectural building blocks, such as hairpins, internal loops, junctions, and pseudoknots (Chastain & Tinoco, 1991). Helical junctions have been shown to be an important motif for establishing global RNA structure, such as in the 5S rRNA (Westhof *et al.*, 1989; Brunel *et al.*, 1991) and active sites of catalytic domains, as in the hammerhead ribozyme (Long & Uhlenbeck, 1993). Several features shared by a number of nucleic acid junction motifs contribute to determine their folded conformation: interactions with polyvalent cations, coaxial stacking of helices, and non-canonical base pairing.

The binding of polyvalent cations to nucleic acid junction motifs is generally accompanied by a discrete folding process. Studies of model DNA four-way junctions that mimic the Holliday junction, the center of homologous and site-specific DNA

recombination (Lilley, 1994; Seeman & Kallenbach, 1994), demonstrated that in a low ionic strength environment and in the absence of divalent metal ions, these junctions have four-fold symmetry, suggesting a pyramidal or square conformation (Clegg *et al.*, 1994). Upon addition of 100 μM $MgCl_2$, the four-way junction adopts an X-shaped conformation in which the four helical arms become coaxially stacked pairs (Cooper & Hagerman, 1987; Duckett *et al.*, 1988; Cooper & Hagerman, 1989; Murchie *et al.*, 1989). RNA four-way junctions, one of which is believed to form the heart of the spliceosome (Steitz, 1992), demonstrate an identical pattern of behavior (Duckett *et al.*, 1995). Similar studies on model DNA three-way junctions, however, showed that they do not undergo an ion-induced conformational change and remain in a symmetrical, Y-shaped conformation (Duckett & Lilley, 1990). Instead, these motifs require the presence of two or more unpaired nucleotides located at the junction to form a cation-stabilized fold (Leontis *et al.*, 1991), a situation that strongly destabilizes the folding of four-way junctions (Duckett & Lilley, 1991). Three-way junction motifs with unpaired nucleotides fold into a well defined structure involving the coaxial stacking of two of the helices in the presence of divalent metal ions, with the third helix unstacked and extended away at an acute angle (Welch *et al.*, 1993) and with differing isomeric forms depending upon the ionic strength and nucleotide sequence at the junction (Welch *et al.*, 1995). In all of these model systems, the effect of divalent ions is to stabilize a conformational change in the junction motif involving coaxial stacking, which is likely to be a recurring theme in all RNA and DNA helical junctions.

Three-way junction motifs in functional RNAs, unlike the aforementioned artificial constructs, often contain nucleotides and non-canonical base pairs that are critical for their folding and function. The 5S rRNA, a structural component of the large ribosomal subunit, contains one three-way junction, loop A, which determines the relative orientation of helices I, II, and V with respect to one another. Extensive phylogenetic analysis and chemical footprinting of this region of the RNA suggest that this junction forms a discrete

structure that is analogous in *X. laevis* and *E. coli* 5S rRNAs in which a proposed base-triple plays an important role (Westhof *et al.*, 1989; Brunel *et al.*, 1991). Solution studies performed on the loop A motif of *Sulfolobus acidocaldarius* 5S rRNA using transient electric birefringence suggest that in the presence of magnesium ions, this RNA adopts a conformation such that helices I and V are colinear, with helix II relatively free to reorient with respect to the other two helices (Shen & Hagerman, 1994). An RNA three-way junction motif that provides the scaffold for a chemical reaction is found in the hammerhead ribozyme. The catalytic core of this ribozyme consists of an internal loop at the intersection of three helices, and contains a number of conserved nucleotides critical for catalytic activity (Long & Uhlenbeck, 1993). In the presence of magnesium, a required cofactor for this RNA enzyme, the hammerhead motif specifically cleaves a substrate RNA oligonucleotide (Scott & Klug, 1996). A number of biochemical, biophysical, and structural studies of the hammerhead have lead to a detailed structural picture of this three-way junction motif. Using native polyacrylamide gel electrophoresis, it was demonstrated that a structural transition occurs in the hammerhead ribozyme upon the addition of polyvalent ions (Bassi *et al.*, 1995), and that the folding of the junction is sensitive to the nucleotide composition in the junction core; mutations in this region disrupt the ability of the RNA to adopt the folded conformation, even in the presence of high divalent metal ion concentrations (Bassi *et al.*, 1996). Thus, the proper folding of the junction is critically dependent upon the nucleotide sequence in the junction element, along with the presence of metal ions. This is clearly revealed in the X-ray crystallographic structure of this ribozyme, in which there are extensive tertiary interactions in the junction core that are consistent with the sequence requirements for hammerhead function (Pley *et al.*, 1994b; Scott *et al.*, 1995b). The recent structure of the P4/P6 domain of the group I intron similarly demonstrates the importance of these features in the J5abc three-way junction (Cate & Doudna, 1996; Cate *et al.*, 1996). From these two structures, a picture is beginning to

emerge of the importance of coaxial stacking, metal ion binding, and non-canonical base pairing in dictating the junction fold.

Like the hammerhead ribozyme, the BS15 rRNA binding site is a three-way junction that contains nucleotides that are conserved and essential for function. In the previous chapter, it was demonstrated that in the absence of divalent metal ions, BS15 was able to bind to this junction and induce a significant conformational change in the helical orientation. Here, the effect of polyvalent ions on the ability of the rRNA three-way junction element to fold in the presence and absence of BS15 is investigated using native polyacrylamide gel electrophoresis and the contribution of the nucleotide sequence at the junction to metal-induced folding.

Folding of Fr1 RNA and mutants in the presence of polyvalent ions

The ability of S15 to enhance the electrophoretic mobility of Fr1 RNA in a native polyacrylamide gel has been interpreted to be the result of a conformational change in the RNA upon protein binding (Batey & Williamson, 1996a; Batey & Williamson, 1996b). To test whether a similar conformational change could be facilitated solely by polyvalent cations, Fr1 RNA, along with three site-specific mutants were subjected to native polyacrylamide gel electrophoresis in the presence of various cations. Two of the point mutants, Fr1(U653Δ) and Fr1(G752C) (see Figure 4.1), are located in the phylogenetically conserved three-way junction element, whereas the third, Fr1(A748Δ), lies inside helix 22.

In a native gel containing EDTA, all four of these RNAs comigrate in the absence of BS15, as shown in Figure 5.1a, indicating that they adopt similar conformations in the absence of polyvalent ions. As observed previously, Fr1 and Fr1(U653Δ) are capable of binding BS15, and display a mobility enhancement characteristic of their specific

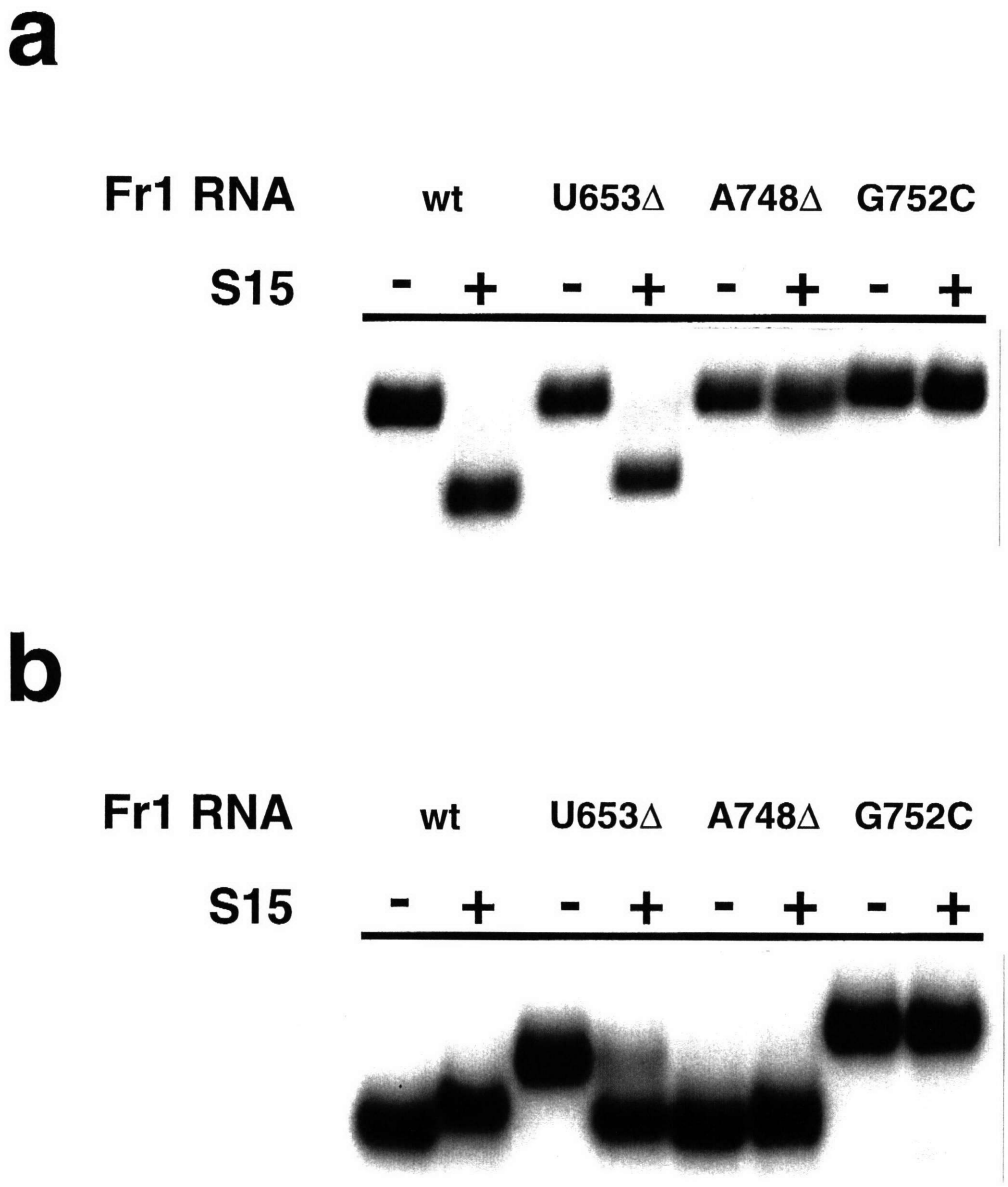


Figure 5.1: Electrophoresis of Fr1 RNA and mutants in 8% nondenaturing polyacrylamide gels containing (a) 1 mM EDTA or (b) 1 mM magnesium chloride in the running buffer, and in the absence and presence of 100 nM BS15.

interaction with the protein. The mobility enhancement of the BS15-Fr1(U653Δ) complex was less than that of the wild type complex, suggesting that this mutation introduces a perturbation in the ability of the RNA to achieve the native conformational change, despite its ability to bind BS15 with wild type affinity. The other two RNAs, as demonstrated previously, are not capable of binding BS15 with high affinity.

This same series of RNAs and protein-RNA complexes demonstrates markedly different mobility patterns when subjected to electrophoresis in the presence of 1 mM magnesium chloride. Fr1 and Fr1(A748Δ), both of which contain the wild type three-way junction element, display equivalent mobilities in a native gel, while the two RNAs containing mutant three-way junctions, Fr1(U653Δ) and Fr1(G752C), display decreased mobility relative to wild type RNA (Figure 5.1b). The mobility retardation of Fr1(U653Δ) is not as severe as Fr1(G752C), suggesting that this RNA is capable of a distorted conformational change around the three-way junction, consistent with its diminished mobility enhancement upon BS15 binding. The presence of magnesium did not alter the ability of Fr1 or the mutants to bind BS15 (*i.e.*, RNA mutants that were incapable of binding S15 in the absence of divalent ions were not rescued by the addition of magnesium), with the only significant difference being that the BS15-Fr1 RNA complex shows a slight retardation with respect to the free RNA in a gel containing magnesium. It appears that this cation plays a similar role to BS15 in the organization of the rRNA three-way junction.

Other polyvalent cations demonstrated similar effects on the mobility of the free and bound RNAs. For instance, calcium chloride, which behaves chemically very similarly to magnesium chloride (Pan *et al.*, 1993), induced exactly the same effect in the mobility patterns. Another polyvalent cation that is well known to stabilize RNA structure is hexamine cobalt complex. Like magnesium and calcium, cobalt(III) forms coordination complexes with six ligands such as H_2O or NH_3, with similar octahedral geometries and

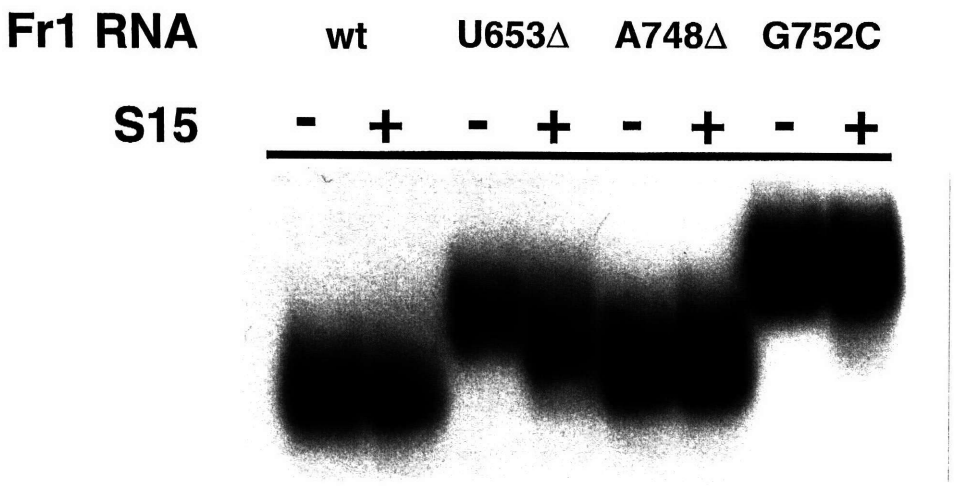


Figure 5.2: Electrophoresis of Fr1 RNA and mutants in an 8% nondenaturing polyacrylamide gel containing 25 μM hexamine cobalt (III) chloride in the gel and running buffer in the absence and presence of 100 nM BS15.

van der Waal radii (Cate & Doudna, 1996) but undergoes ligand exchange extremely slowly (Pan *et al.*, 1993) and has particular affinity for nitrogen donors such as $-NH_2$ and the N7 of purines (Cotton *et al.*, 1987). Despite the fully occupied, kinetically inert coordination sphere, hexamine cobalt is still capable of inducing the same conformational change in the RNA (Figure 5.2), suggesting that the conformational change can be accommodated by a variety of hexa-coordinated metal ions. The organic polyamine spermidine is also capable of inducing a moderate conformational change about the three-way junction; at concentrations of 200 μM and 1 mM, spermidine causes a mobility enhancement in the wild type RNA that is similar to that observed for the metal ions, but to a lesser degree (Figure 5.3). Thus, in the presence of spermidine, the BS15-Fr1 RNA complex has a slightly greater mobility to that of free Fr1, although not nearly as great as that observed in EDTA. This strongly indicates that the three-way junction does not contain a specific metal-binding pocket. Instead, the three-way junction is capable of achieving the folded conformation with the aid of diverse ligands such as metal ions, organic polyamines, and the protein S15.

Modification interference of the RNA conformational change

Since mutations in the three-way junction are capable of affecting the mobility of Fr1 RNA in a native polyacrylamide gel, the modification interference assay can be utilized to elucidate functional groups in the RNA that are important for the divalent ion-induced conformational change. This experiment, however, had to be performed slightly differently from the S15 modification interference experiments, since no single magnesium concentration in the gel yielded separate "free" and "bound" bands, as observed in experiments using protein. Instead, the modified RNA was subjected to electrophoresis on two separate polyacrylamide gels, one containing EDTA and one containing 1 mM

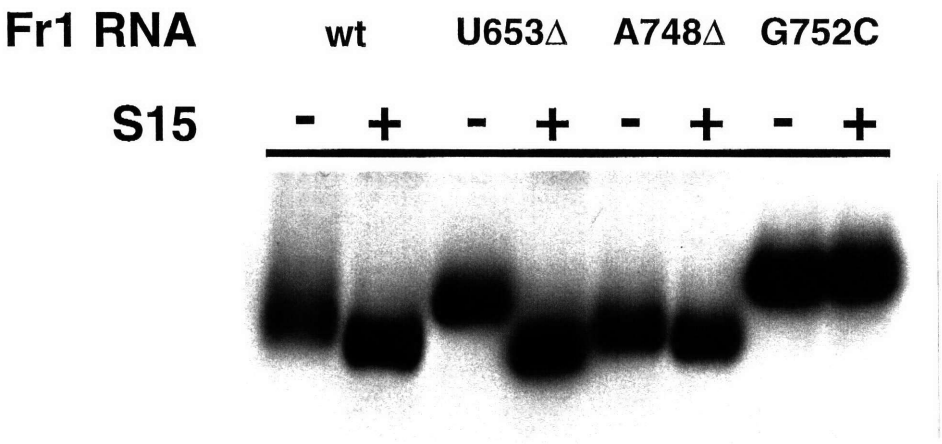


Figure 5.3: Electrophoresis of Fr1 RNA and mutants in an 8% nondenaturing polyacrylamide gel containing 200 μM spermidine chloride in the gel and running buffer in the absence and presence of 100 nM BS15.

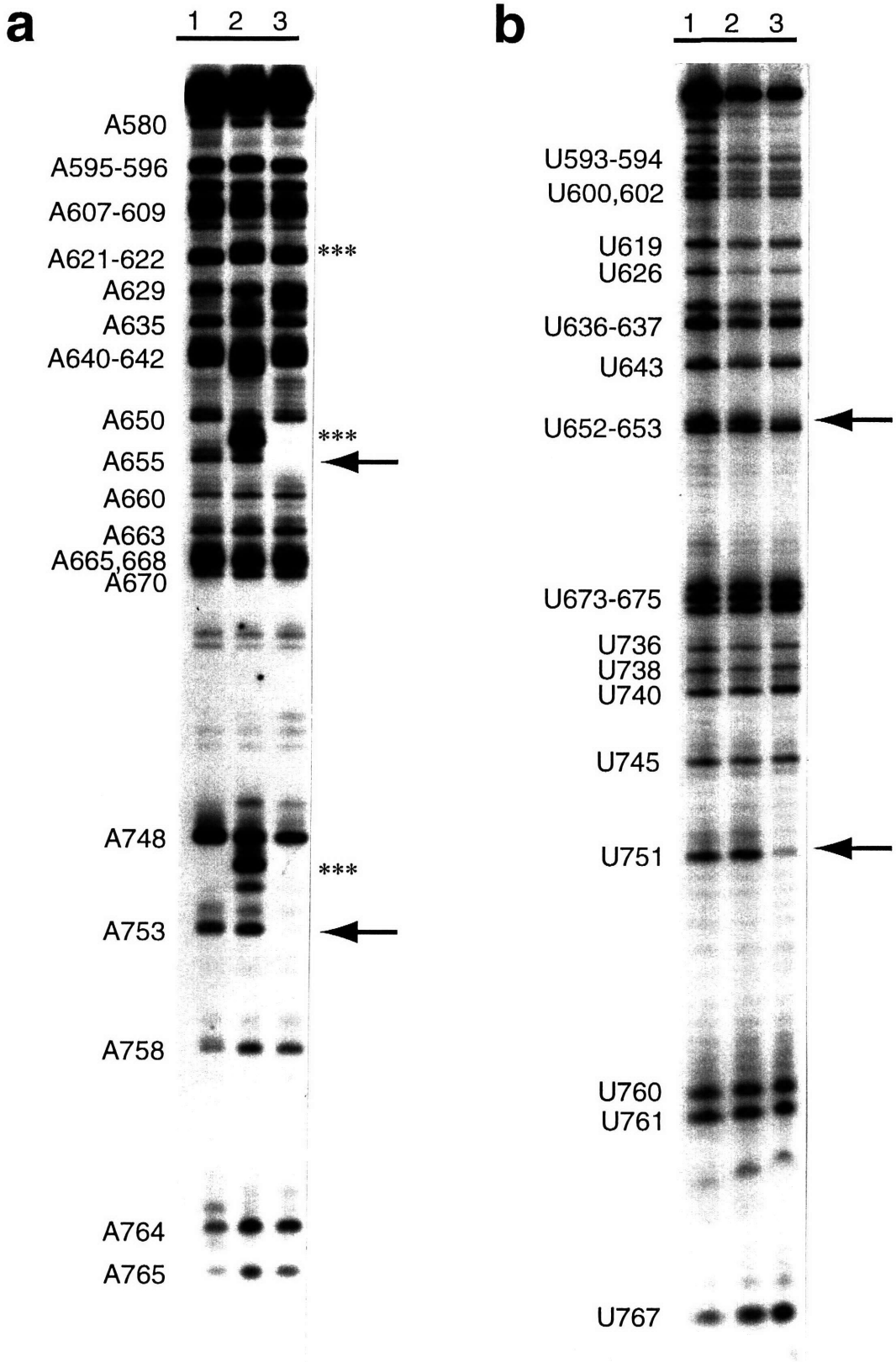
magnesium chloride, with the EDTA gel serving as a control for modifications that do not specifically interfere with the cation-induced RNA conformational change. From each of these gels, the major RNA band was excised and subjected to strand scission and sequencing, as described previously.

Modification interference experiments were performed on 3' end-labeled Fr4 RNA (Figure 3.3) using DEPC and aqueous hydrazine. Fr4 was chosen for these studies because it contains the three-way junction element, displays an enhanced mobility upon organization of the RNA by specific recognition of BS15 (Figure 3.4), and since it lacks helix 23, reading the sequence of the entire RNA becomes significantly easier. Modifications of Fr4 with DEPC (Figure 5.4a) and aqueous hydrazine (Figure 5.4b) reveal four nucleotides which are involved in the folding of the RNA about the three-way junction motif. DEPC modifications of A655 and A753, both located at the junction, strongly interfere with magnesium dependent RNA folding; these adenosines are also implicated in recognition of the RNA by BS15 (Figure 4.3). However, modifications at A663 and A668, which are located in the middle of helix 22 near the internal loop and also implicated in protein binding, do not interfere with RNA folding. Likewise, the hydrazine-induced depyrimidinations at U652 and U751, which are located near the three-way junction, are important for both protein binding and magnesium dependent folding. Modification of U740, another important determinant for BS15 binding that is located well outside the junction element in helix 22, does not interfere with the ability of the RNA to fold. Thus, nucleotides important for the magnesium dependent folding of the RNA are localized to the three-way junction element and overlap with nucleotides in this region that are important for protein binding.

The relationship between RNA folding and protein binding

It has been previously demonstrated that determinants for recognition of 16S rRNA by the primary ribosomal binding protein S15 from *Bacillus stearothermophilus* lie in two

Figure 5.4 (next page): Modification interference assays of Mg^{2+} binding to 3' end-labeled Fr4 RNA using (a) DEPC and (b) aqueous hydrazine as modification agents. Individual lanes are: (1) modified RNA not previously subjected to gel electrophoresis, (2) RNA previously subjected to gel electrophoresis in the presence of EDTA, and (3) RNA previously subjected to gel electrophoresis in the presence of 1 mM magnesium chloride. Arrows denote nucleotide positions whose modification interferes with magnesium dependent folding, and asterisks denote sites of non-specific RNA hydrolysis.


a
1 2 3
A580
A595-596
A607-609
A621-622
A629
A635
A640-642
A650
A655
A660
A663
A665,668
A670
A748
A753
A758
A764
A765
***
***
***
b
1 2 3
U593-594
U600,602
U619
U626
U636-637
U643
U652-653
U673-675
U736
U738
U740
U745
U751
U760
U761
U767

portions of helix 22, near an internal loop motif and a three-way junction (Batey & Williamson, 1996b). The junction, a site containing four extremely highly phylogenetically conserved nucleotides, has been proposed to undergo a conformational change upon protein binding, including the formation of several non-canonical base pairs (Batey & Williamson, 1996b). One question that arose from those studies was whether folding of this junction motif required specific functional groups supplied by S15 or that the conserved nucleotides in the junction were the principle participants. Native gel electrophoresis of Fr1 RNA and several point mutants in the presence of chemically different polyvalent ions clearly indicates that the RNA is capable of a conformational change about the three-way junction without the involvement of S15. Thus, it appears that S15 folds the junction in the absence of other polyvalent ions by providing appropriately placed arginine or lysine residues, and that RNA folding is likely to be accomplished by relieving high negative charge density of the phosphate backbone in the folded conformation. The interrelationship of nucleotides important for protein binding and RNA folding is clearly demonstrated with the same approach used previously to elucidate the protein binding determinants in 16S rRNA. The phylogenetically conserved three-way junction element (Figure 4.10a) is an independently folding element with the 16S rRNA with the phylogenetically conserved nucleotides primarily fulfilling a structural role, with the protein specifically recognizing the folded RNA. However, the site-specific mutants demonstrate that the nucleotides that affect protein binding and RNA folding in the junction do not completely overlap. While the G752C mutant destroys both folding and binding, the U653Δ only significantly affects RNA folding.

The ability of the rRNA three-way junction to fold independently of S15 suggests that the mechanism of protein binding may not be the simple bimolecular association illustrated in Figure 4.12. Recently, it was demonstrated that the folding of the bI5 group I intron was facilitated by the interaction of the protein CBP2 (Weeks & Cech, 1995). In the

presence of low concentrations of magnesium (7 mM), the intron alone is in dynamic equilibrium between a form in which only the secondary structure has formed and one in which the catalytic core has formed. CBP2 binds to the transiently formed tertiary structure of the catalytic core, stabilizing it and allowing it to finish folding into an active enzyme (Weeks & Cech, 1996). The instability of the tertiary structure of the intron's core can be alternatively overcome with high concentrations of magnesium (40 mM). Thus, the role of the protein is to stabilize high-order RNA structure that is inherently unstable, a common problem for many large, functional RNAs. S15 could be binding 16S rRNA via a similar mechanism, in which the three-way junction is in a dynamic equilibrium between the unfolded and folded form, and S15 interacts specifically only with the folded three-way junction element (scheme shown in Figure 5.5). In the absence of divalent ions, this equilibrium lies predominantly towards the unfolded form of the RNA, while in the presence of divalent ions, the equilibrium lies towards the folded form. There is significant evidence that this is the mechanism of binding of S15 to mRNA to regulate its own translation. The translational regulatory site of the S15 mRNA is in dynamic equilibrium between stem-loop and pseudoknot forms, with S15 binding only specifically to the pseudoknot (Portier *et al.*, 1990c; Phillipe *et al.*, 1993). Mutations which stabilize the pseudoknot increase autoregulation of the gene *in vivo* (Portier *et al.*, 1990c; Benard *et al.*, 1994) and in vitro binding affinity (Philippe *et al.*, 1994). The similarities in the mRNA and rRNA S15 recognition sites (Batey & Williamson, 1996b), may also be reflected in their mechanism of S15 binding.

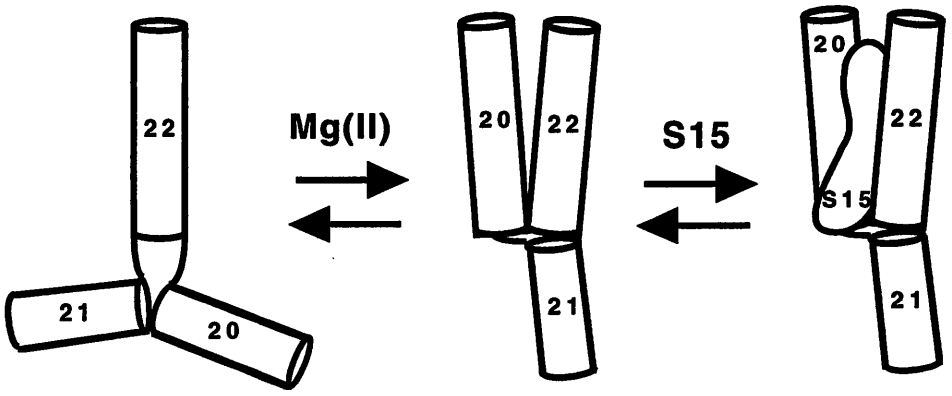


Figure 5.5: Alternative model for BS15 binding to 16S rRNA based upon the ability of the three-way junction element to self-organize, whose folding is stabilized by the presence of multivalent cations.

Chapter 6: NMR studies of the BS15-rRNA interaction

In previous chapters, it was demonstrated that BS15 was capable of specifically interacting with a minimal RNA element of 61 nucleotides (Fr16 RNA), which was a critical step towards the development of a model BS15-rRNA complex that could be utilized for structural studies. Unfortunately, a minimized protein-RNA complex that is capable of being characterized using biochemical means will not simply translate into a system that ready to study using NMR spectroscopy. In developing an NMR-amenable complex, a number of issues need to be addressed in a phase of structural studies referred to as "sample conditioning", which is particularly critical since the minimal BS15-rRNA complex is pushing the classical ~30 kDa NMR size limitation. In this chapter, the steps that have been taken to address these issues with this complex are discussed. Also, an effort to develop a minimal RNA three-way junction element for use in structural studies will be detailed.

NMR spectroscopy of the BS15-rRNA complex

Quality of RNA and protein utilized for NMR studies

All of the RNAs that have been characterized in detail by NMR spectroscopy have been less than 40 nucleotides in length. For these small RNAs, the method used to reproducibly obtain large quantities (approximately 1 micromole of RNA is needed per NMR sample) of high quality RNA is single-stranded template directed transcription using T7 RNA polymerase, as detailed in Appendix 1. However, when this method was applied to the synthesis of a 65 nucleotide RNA (similar to Fr15, Figure 3.3), the resulting product

yielded poor NMR spectra with extremely broad linewidths and was reduced in its ability to bind BS15. The fundamental problem was in the synthesis and purification of suitable quantities of the 82mer DNA oligonucleotides required as a template for this RNA. First, the laboratory DNA synthesizer (Applied Biosystems PCR Mate 391) could not reproducibly achieve the >99% coupling efficiencies in a 1 μmole-scale synthesis necessary to recover high quantities of oligonucleotide (typically, a coupling efficiency of ~97.5% was observed, which led to a 13% yield ($.975^{81}$) for the overall synthesis). Second, DNA oligonucleotides are typically purified by electrophoresis in denaturing polyacrylamide gels. For longer oligonucleotides, achieving nucleotide resolution on preparative gels is impractical, which can translate into inhomogeneity at the 3' end of the RNA product. Finally, at the time these experiments were being performed, dichloroacetic acid was used as the detritylating agent during synthesis, rather than the standard trichloroacetic acid. This leads to incomplete detritylation at each step of the synthesis, which manifests itself as internal deletions in the resulting oligonucleotide (Treiber & Williamson, 1995). These problems were circumvented by utilizing plasmids as templates for T7 RNA polymerase directed transcription. This approach was not initially attempted since 2-3 milligrams of plasmid template were required for 60 ml of transcription (sufficient to yield one NMR sample).

For biochemical studies of the BS15-rRNA interaction, sufficient purity had been obtained after the first two chromatographic steps of the purification procedure detailed in Appendix 1. However, despite been purified to >99% homogeneity, the protein preparations used for the biochemical studies were unsuitable for use in NMR spectroscopy. While this protein could bind RNA in the NMR tube, and remained soluble at 1.5 mM concentrations, the protein-RNA complex lacked long-term stability. At 45 °C, the temperature at which many spectra were obtained (detailed below), the protein-RNA complex slowly aggregated and precipitated over the course of 1-2 weeks, which led to a dramatic decrease in the quality of the data. This problem was ameliorated by the addition

of two subsequent chromatographic steps in the protein purification procedure, which removed a lot of the microinhomogeneities in the BS15 preparation (note the removal of protein impurities between lanes 5 and 7 in Figure A1.3). The fully purified protein displayed no solubility or precipitation problems, which is critical for the ability to acquire high quality data throughout the lifetime of an NMR sample.

Formation of the BS15-rRNA complex in the NMR tube

For NMR studies, a minimal BS15-rRNA complex was formed by direct titration of a 600 µl sample of RNA in a 90%H_2O/10%D_2O in a buffer containing 10 mM potassium phosphate, pH 6.5, and 25 mM potassium chloride at 25 °C, with one equivalent of BS15 while monitoring the imino proton NMR spectrum of the RNA. In a typical titration, a 1.2 mM sample of Fr15 RNA was titrated by addition of four 26 µl aliquots of a 4.8 mM solution of BS15 in ddH_2O. The imino proton region of the RNA spectrum is shown in Figure 6.1 at each point during the titration. The clear change in the imino spectrum upon addition of a stoichiometric amount of protein demonstrates the formation of a single, specific complex. The fact that the linewidths of the imino peaks do not broaden significantly during the titration indicates that the complex is soluble at a concentration above 1 mM and behaves as a monomeric species. Furthermore, the observation that at a protein:RNA ratio of 0.5:1 two sets of peaks were observed demonstrates that the complex is in slow exchange on the NMR timescale at 25 °C. RNA fragments Fr11, Fr12, and Fr16 were also titrated with BS15 in the same manner and display the same characteristic changes in the imino spectrum (data not shown). Further titration of RNA beyond a 1:1 stoichiometry produce no further changes in the imino spectrum, verifying the specificity and stoichiometry of the complex. The preparation of the protein-RNA complex via NMR titration was necessary to prevent overtitration of the

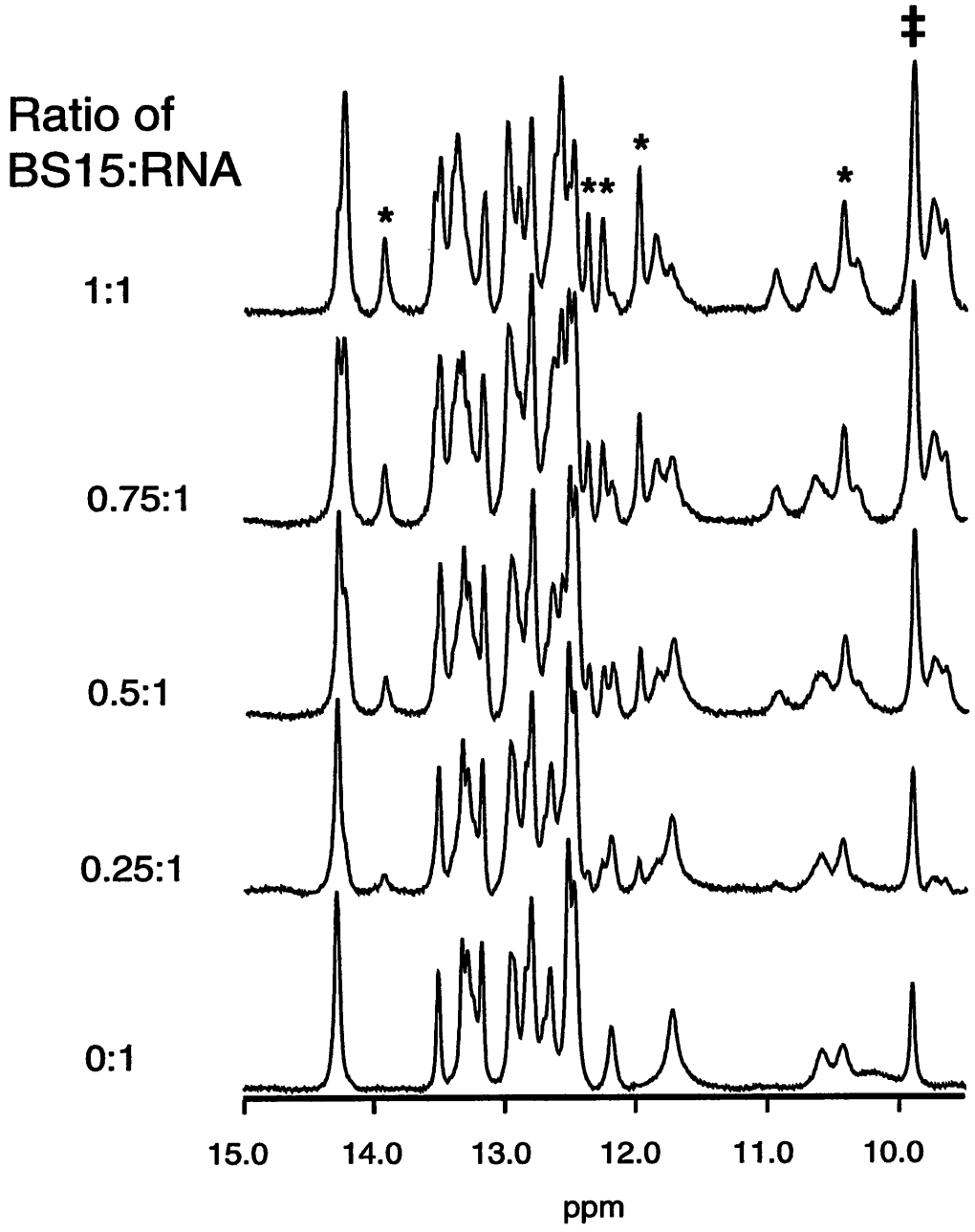


Figure 6.1: Imino proton NMR spectrum of Fr15 RNA and several points during its titration with BS15, monitoring complex formation at 25 °C. Asterisks indicate peaks which appear upon titration of the protein into the RNA sample, and the dagger denotes the peak characteristic of the guanosine imino proton in the UUCG tetraloop.

RNA with protein, since excess protein in the sample invariably precipitates over time, diminishing the quality of the sample.

Optimization of conditions for NMR spectroscopy

Although the imino spectrum of the BS15-Fr15 complex showed promise for further NMR analysis, 2D-NOESY and DQF-COSY spectra of the complex in 99.996% D_2O under the same conditions used for the titration exhibited severely broadened linewidths, which would make subsequent steps in structural characterization difficult. The primary strategies for overcoming these problems are screening pH, temperature, sample concentration, ionic strength, and buffer of the solution and even mutants of the macromolecule or complex of interest for improvements in spectral quality (Wagner, 1993). The parameter that has the greatest impact upon the spectral quality of this complex is temperature. As discussed in Chapter 2, the correlation time of a molecule in solution, an important determinant of spectral linewidths, is inversely proportional to the temperature. Thus, as the temperature of the sample increases, the observed linewidths should decrease. Also, since intrinsic viscosity of the solvent decreases as a function of temperature, this also favors shorter correlation times and narrower linewidths. Since *B. stearothermophilus* is a moderate thermophile whose optimal growth occurs at 60 °C, it was reasoned that the BS15-Fr15 complex would be resistant to dissociation at elevated temperatures. The similar BS15-Fr11 and -Fr12 complexes have half-lives of dissociation of at least five minutes at 45 °C (Table 3.3), and thus it is not likely that an intermediate exchange regime will be entered by observing the complex at elevated temperatures. Figure 6.2 shows the effect of increasing the temperature from 25 °C to 45 °C on the RNA region of the 1D NMR spectrum of the BS15-Fr15 complex. The linewidths of the RNA base and ribose H1' proton resonances clearly decrease as the temperature increases, a trend that is also observed in experiments which measure the T_2 relaxation time (Table

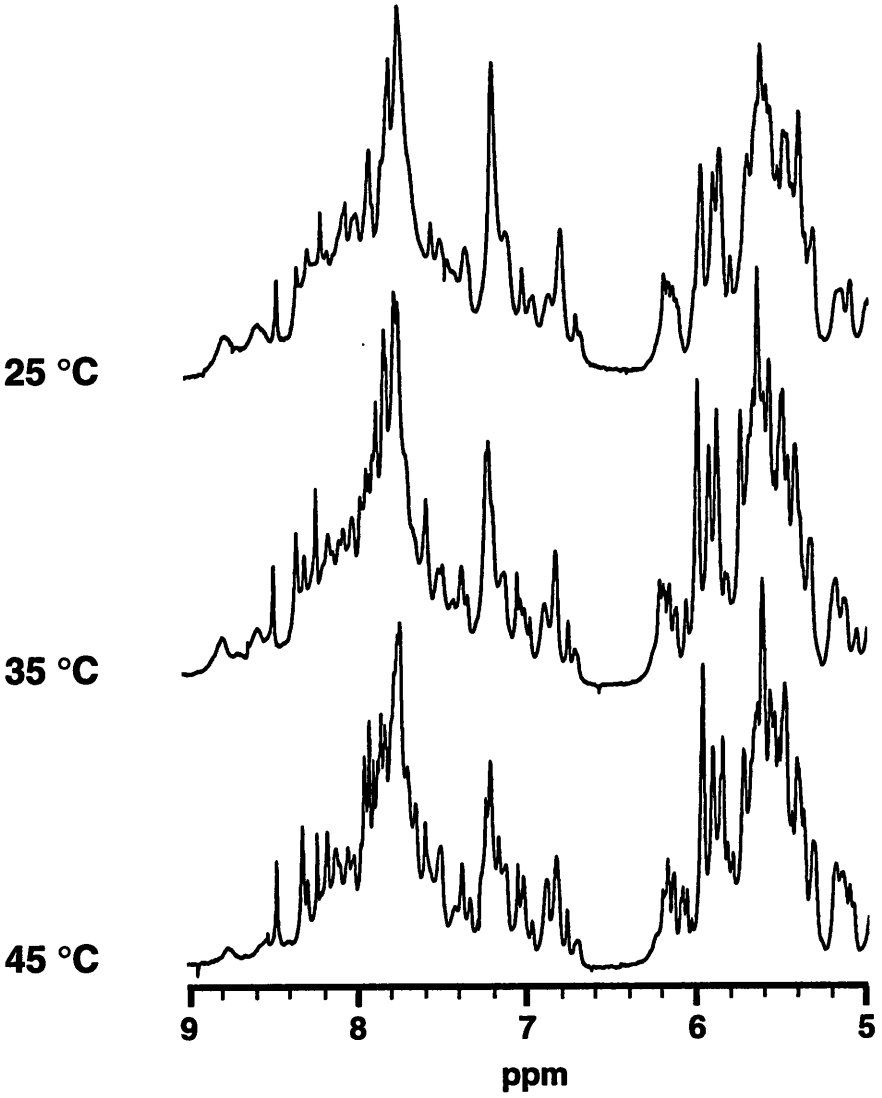


Figure 6.2: Temperature dependence of the 1D-NMR spectrum of the BS15-Fr15 complex. Each spectrum was acquired and processed identically, so as to be able to make direct comparisons. As the temperature of the sample is increased from 25 °C to 45 °C, the linewidths of the RNA proton resonances (these spectra highlight the RNA aromatic and anomeric region) sharpen considerably.

Table 6.1: Transverse relaxation times (T_2) for the BS15-Fr15 complex as a function of temperature.

	$\langle T_2 \rangle^a$, seconds		
Temperature	7.6-8.0 ppm (H6/H8/H2)	6.0-6.5 ppm (H1'/H5)	0.8-1.3 ppm (methyl, protein)
25 °C	0.0146	0.0175	0.0082
35 °C	0.0183	0.0222	0.0107
45 °C	0.0220	0.0249	0.0131
55 °C	0.0227	0.0252	0.0140

[a]The average T_2 relaxation time, calculated by averaging all of the proton resonances in the defined regions because of spectral overlap.

6.1). In the case of both protein and RNA resonances, the relaxation time increases significantly between 25 °C and 45 °C, and then begins to level off at 55 °C. These experiments, along with 2D DQF-COSY spectra (data not shown), clearly indicate that optimal spectral quality is achieved between 45 °C and 55 °C while still maintaining a specific protein-RNA complex that is in slow exchange on the NMR timescale.

While increasing the temperature significantly improved the spectral linewidths, the RNA in the sample suffered from a dramatically increased rate of nonspecific hydrolysis. This problem manifested itself as new peaks growing into the DQF-COSY spectrum with time, and could also be monitored by subjecting a small amount of the sample to electrophoresis on a 12% denaturing polyacrylamide gel. Although there were a few hot spots of degradation (such as at the G_{752}pA step), hydrolysis occurred throughout the RNA. Furthermore, Fr15 RNA is stable for prolonged periods (> 1 month) at 45 °C with less than 5% degradation in the absence of BS15, but upon introduction of a stoichiometric quantity of protein, degradation started to occur rapidly. Treatment of the BS15 sample with pronase (a nonspecific cocktail of proteases) or DEPC (which modifies histidine, a common component of the active site of RNases) completely suppressed the ability of the BS15 preparation to hydrolyze Fr15 RNA. Also, addition of 10 μg/μl heparin, which is a potent inhibitor of protein-RNA interactions, suppresses the hydrolysis problem. Thus, it appeared that the BS15 preparation was contaminated with a protein RNase. However, further purification of BS15, including size exclusion chromatography and native ion exchange chromatography (DEAE and CM), did not diminish the nonspecific hydrolysis; the RNase activity seemed to copurify with BS15 in every purification step tested, suggesting that the hydrolysis activity could be associated with BS15 itself. To quickly assay degradation activity, different stoichiometric amounts of BS15 was titrated into a 20 μM sample of Fr15 (containing a trace quantity of 3' end-labeled RNA) and incubated at 45 °C for 4 days. A small aliquot of each reaction was electrophoresed on a 12% denaturing

acrylamide gel, as shown in Figure 6.3. The illustrated gel clearly demonstrates that as the concentration of BS15 in the reaction increases, the hydrolysis of RNA increases, to be expected if the protein preparation is contaminated with an RNase activity. Unexpectedly, upon overtitration of Fr15 RNA with BS15 (beyond a 1:1 BS15:Fr15 RNA stoichiometric ratio), the pattern of RNA hydrolysis changes, which should not occur if the contaminant was a trace RNase impurity. Instead, this assay suggests that the hydrolysis activity may be associated, in part, with BS15 itself. By screening for conditions under which this problem was minimized, it was determined that two factors led to minimal RNA degradation at 45 °C. The first, as suggested by the experiment described above, is to leave Fr15 RNA slightly undertitrated (a 0.9:1.0 BS15:Fr15 ratio is optimal). This approach was utilized in the studies of the U1A:U1 snRNA interaction by NMR, which suffered from a similar problem (in this case, the stoichiometry of the complex studied by NMR was 0.7:1.0 U1A protein:U1 snRNA) (Howe *et al.*, 1994). The other parameter which strongly affects the rate of RNA degradation at 45 °C is the ionic strength. It was empirically determined that very low ionic strength conditions effectively suppressed this problem; in a buffer containing solely 10 mM potassium phosphate, pH 6.5, the RNA was reasonably stable over the period of several weeks (<10% degradation of the RNA) in the presence of protein. Again, this strategy has been employed in the studies of other protein:RNA complexes by NMR (Howe *et al.*, 1994; Gubser & Varani, 1996). Optimal homonuclear NMR spectra were obtained using a buffer composed of 10 mM potassium phosphate and 25-50 mM potassium chloride (spectra of the complex could be obtained without the presence of KCl, but the spectral quality was slightly diminished compared to that of spectra obtained in presence of 25 mM KCl).

The other parameter that was explored was the pH of the NMR sample. The pH of a sample has the greatest impact upon the quality of the exchangeable proton spectra, since the rate of proton exchange with bulk solvent, k_{intr}, is a function of pH (Wüthrich, 1986). For instance, the exchange rate of protein backbone amide protons reaches a minimum at

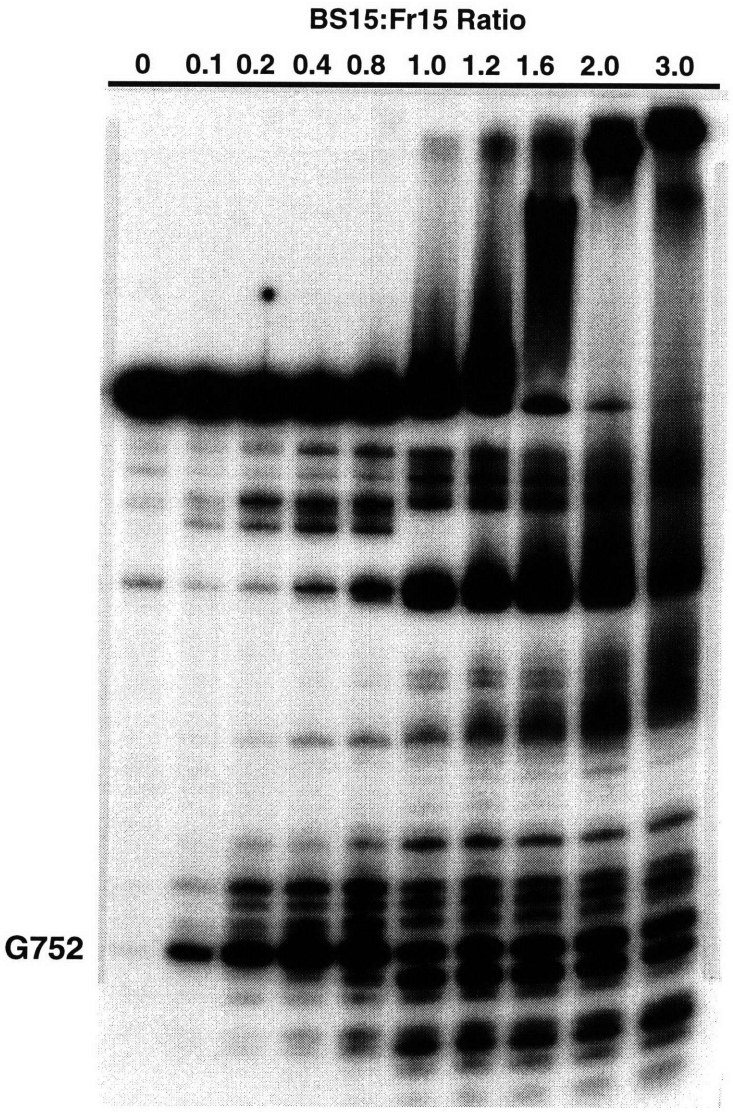


Figure 6.3: Degradation of 3' end-labeled Fr15 RNA in the presence of purified BS15 at various protein:RNA stoichiometric ratios after 4 days at 45 °C. Note that when the protein:RNA ratio is at or exceeds 1:1, the pattern of degradation abruptly changes, suggesting that unbound BS15 might contribute to degradation. The primary site of hydrolysis below a 1:1 protein:RNA ratio is at the G_{752}pA step, while Fr15 in the absence of BS15 remains stable to hydrolysis over the time period tested.

pH ~3, and nucleic acid imino protons reach an exchange rate minimum around pH ~5. Since observation of labile protons by NMR is most feasible when exchange with bulk solvent is slow on the NMR timescale, it was desirable to lower the pH as much as possible. By observing the 1D-imino spectrum of the complex as a function of pH (a 1 mM BS15-Fr15 sample in 10 mM phosphate buffer, pH 7.0, was titrated with 0.1 M HCl down to pH 5.0 in 0.3 pH unit steps), it was determined that the lowest pH which demonstrated clear complex formation was pH 5.7 (data not shown). Thus, to error on the side of caution, pH 6.0 was chosen as the optimal pH.

Under optimal conditions, high quality homonuclear NMR spectra of BS15 complexed with Fr11, Fr12, Fr15, and Fr16 could be obtained. The spectra of the BS15-Fr15 complex appeared to have the highest spectral quality, and thus all subsequent studies have focused on this complex. The 2D-DQF-COSY of the pyrimidine H5-H6 region of Fr15 RNA and the BS15-Fr15 complex are shown in Figure 6.4. Upon complexation with a stoichiometric amount of protein, a number of H5-H6 crosspeaks undergo a large change in chemical shift, consistent with a significant change of the chemical environment of multiple pyrimidine bases. A NOESY spectrum (200 millisecond mixing time) taken at 750 MHz of the BS15-Fr15 complex is shown in Figure 6.5, along with an expansion of the regions of the spectrum where crosspeaks between RNA aromatic and anomeric protons (Figure 6.5b) and between RNA aromatic/protein aromatic and protein methyl protons (Figure 6.5c). Several features of these spectra are indicative that this complex is well-behaved under NMR conditions. In the RNA aromatic-anomeric region, the crosspeaks show a high degree of dispersion with relatively little spectral crowding despite the fact that >150 peaks are expected in this region, which is important because this is a critical region for making RNA sequence specific resonance assignments, and these peaks are fairly sharp despite the 31 kDa molecular weight of this complex. In the aromatic-methyl region of this spectrum, there are numerous crosspeaks between the tyrosine and phenylalanine aromatic protons and the methylene and methyl protons of the protein, indicative of a compact

Figure 6.4 (next page): DQF-COSY spectra of Fr15 RNA taken at 45 °C (a) in the absence of BS15, (b) half-titrated with BS15 (0.5:1.0 BS15:Fr15 RNA), and (c) fully titrated to form a ~0.95:1 BS15:Fr15 RNA complex. In the spectrum of the half-titrated complex, there is clearly two sets of peaks, one corresponding to free RNA and one corresponding to bound RNA, a clear indication that the complex is in slow exchange. Furthermore, the dispersion of the H5-H6 crosspeaks has increased upon binding protein, demonstrating that the presence of BS15 actually helps resolve many of the resonances which are crowded in the free RNA.

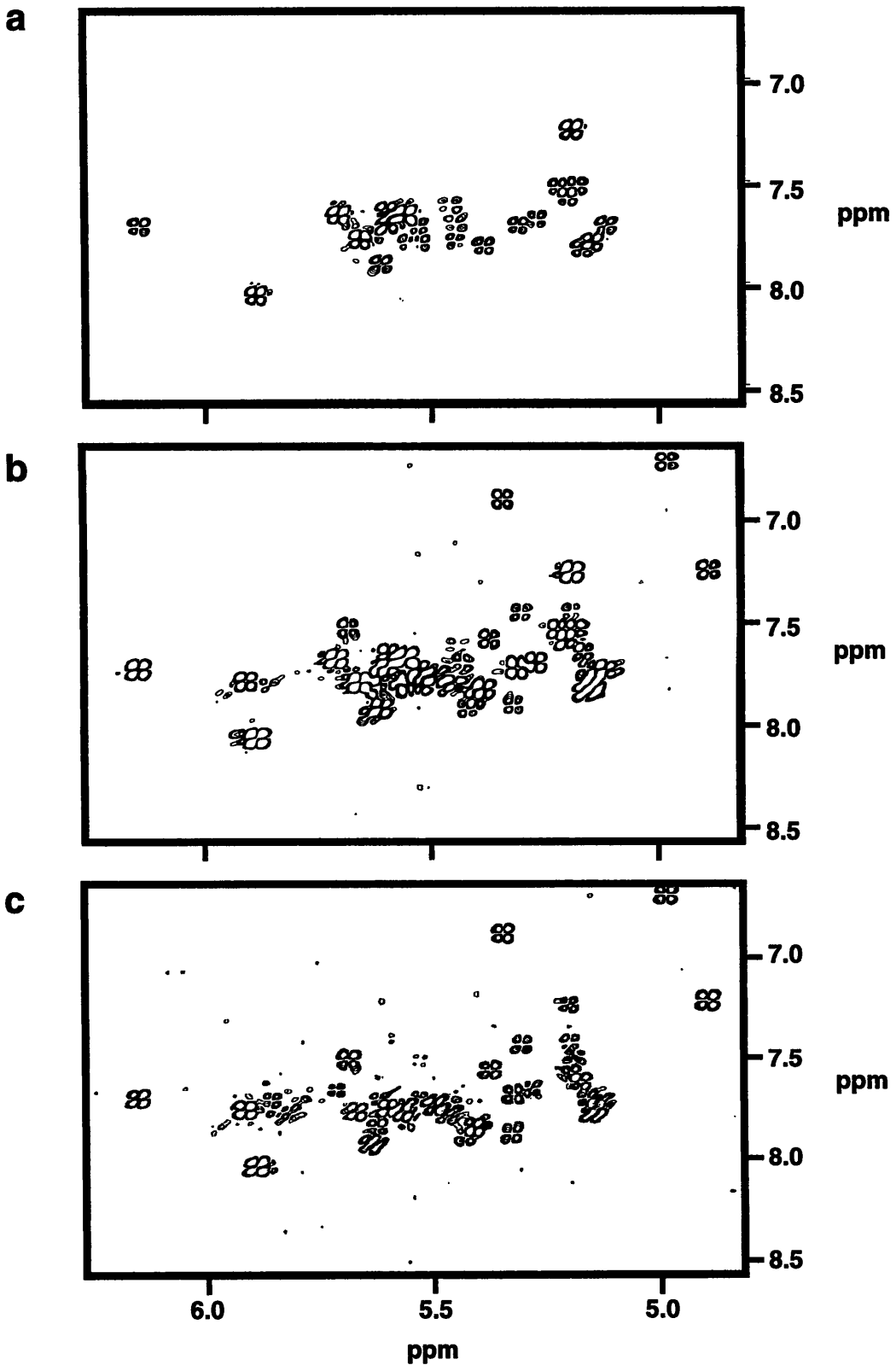


Figure 6.4

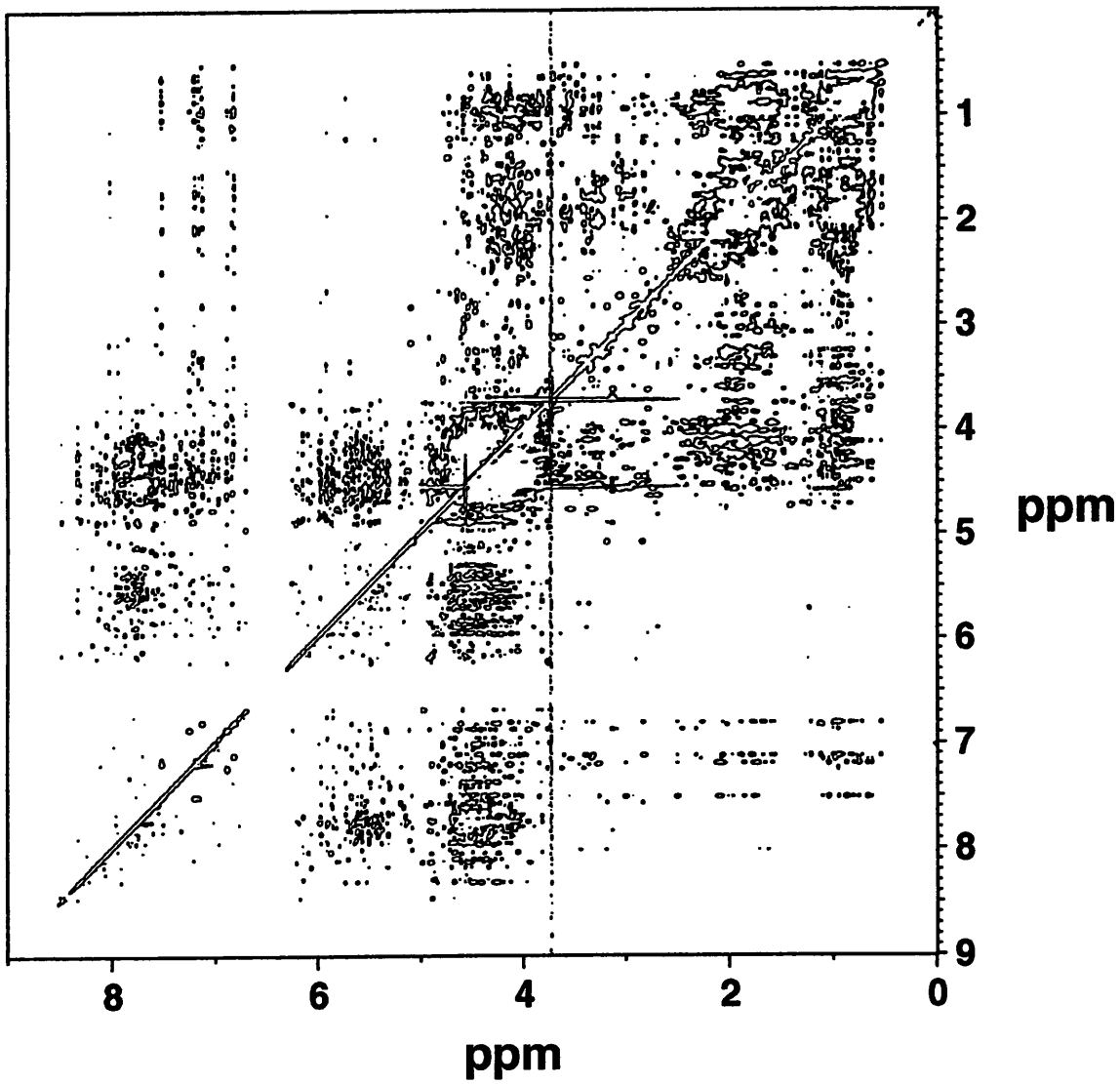


Figure 6.5: (a) 200 ms mixing time NOESY of the BS15-Fr15 RNA complex acquired at 750 MHz at 45 °C. (b), (c) Expansions of two interesting regions of this spectrum are displayed on the following pages.

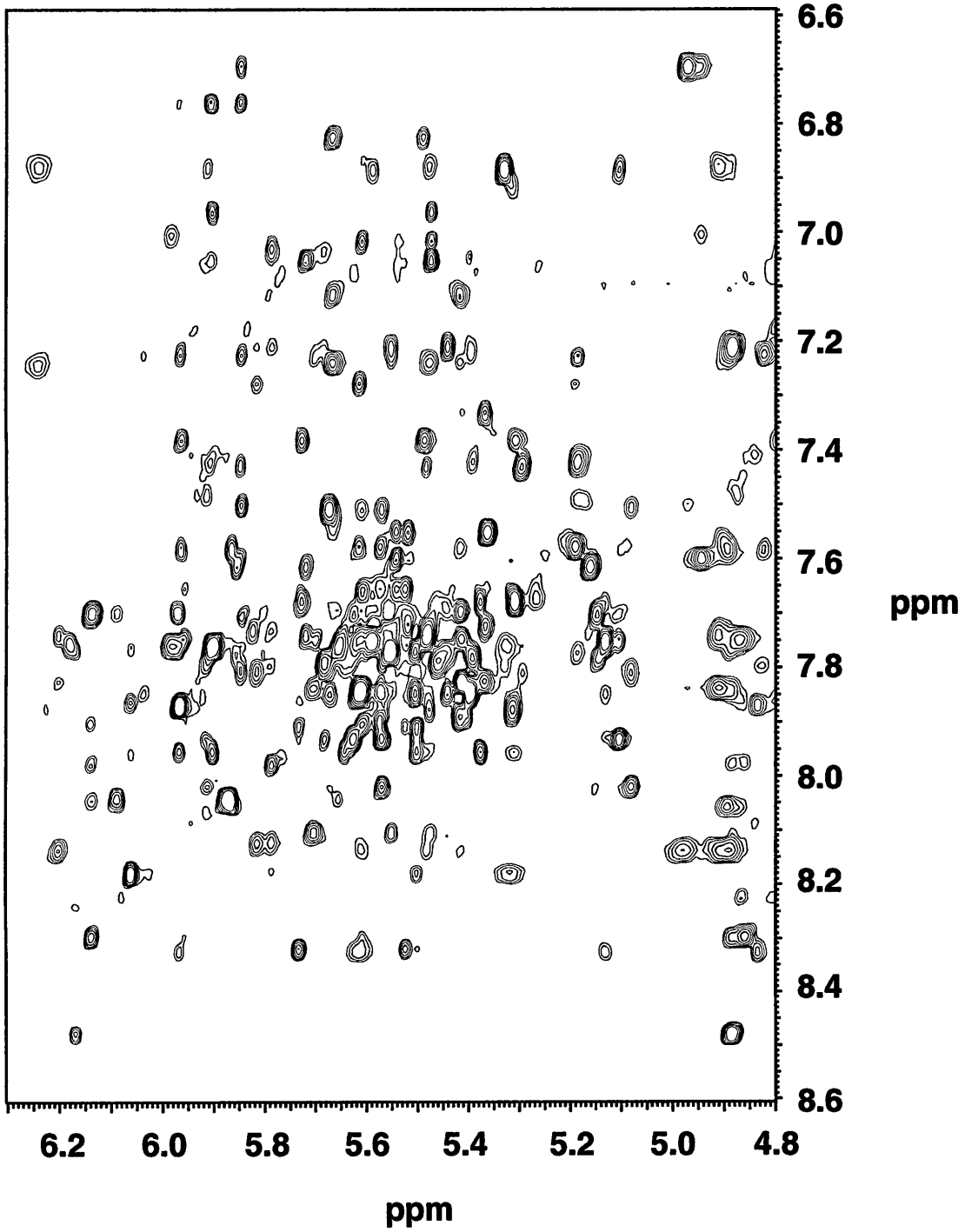


Figure 6.5, (b)

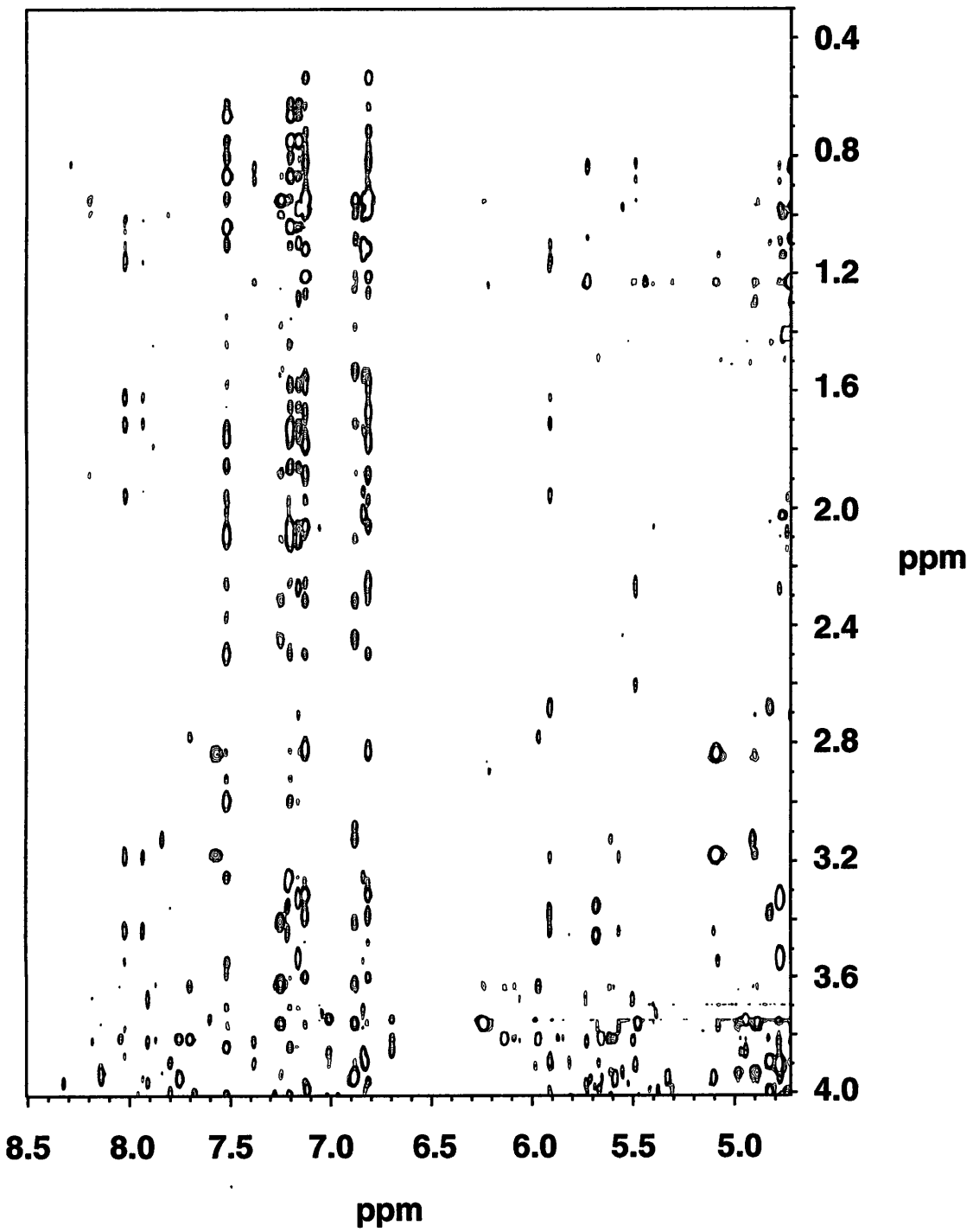


Figure 6.5, (c)

hydrophobic core in the protein. Also, in this region there are a number of crosspeaks between the RNA H1'/H5 region of the spectrum and the protein methylene/methyl region, many of which are likely to represent intermolecular contacts.

Heteronuclear spectroscopy of the BS15-Fr15 complex

The next step in exploring the feasibility of detailed structural characterization of the BS15-Fr15 complex using NMR is to determine whether sequence specific assignments could be made using heteronuclear experiments. The recent development of methods to isotopically label RNA with the NMR-observable heteronuclei ^{13}C and ^{15}N (Batey *et al.*, 1992; Nikonowicz *et al.*, 1992; Batey *et al.*, 1995) and efficient methods for sequence specific assignments using heteronuclear experiments (Nikonowicz & Pardi, 1993) have overcome many of the problems associated with spectral overlap in large RNA molecules. Even though the 2D-NMR spectra of the BS15-Fr15 complex show many well resolved crosspeaks, heteronuclear 3D- and 4D-NMR experiments will be required to completely assign both the protein and the RNA. To this end, a Fr15 RNA was synthesized in which only the guanosine nucleotides are uniformly labeled with ^{13}C (referred to as (^{13}C-G)Fr15). In Figure 6.5, the HSQC spectrum of the BS15-(^{13}C-G)Fr15 complex is shown. Most notably, the resonances of the ribose H2', H3', H4', and H5'/5", which are crowded into a single cluster in the proton dimension are well resolved into three clusters in the carbon dimension based upon their carbon chemical shift. Similarly, complexes labeled with the other three nucleotides individually gave well resolved 2D-HSQC spectra. However, when 3D-NOESY-HSQC, HCCH-TOCSY, and HCCH-COSY experiments were attempted, all of these experiments failed to yield quality spectra. While the NOESY-HSQC spectrum displayed some crosspeaks, they were all severely broadened, and the HCCH-TOCSY and HCCH-COSY spectra failed to display any peaks at all. This is most

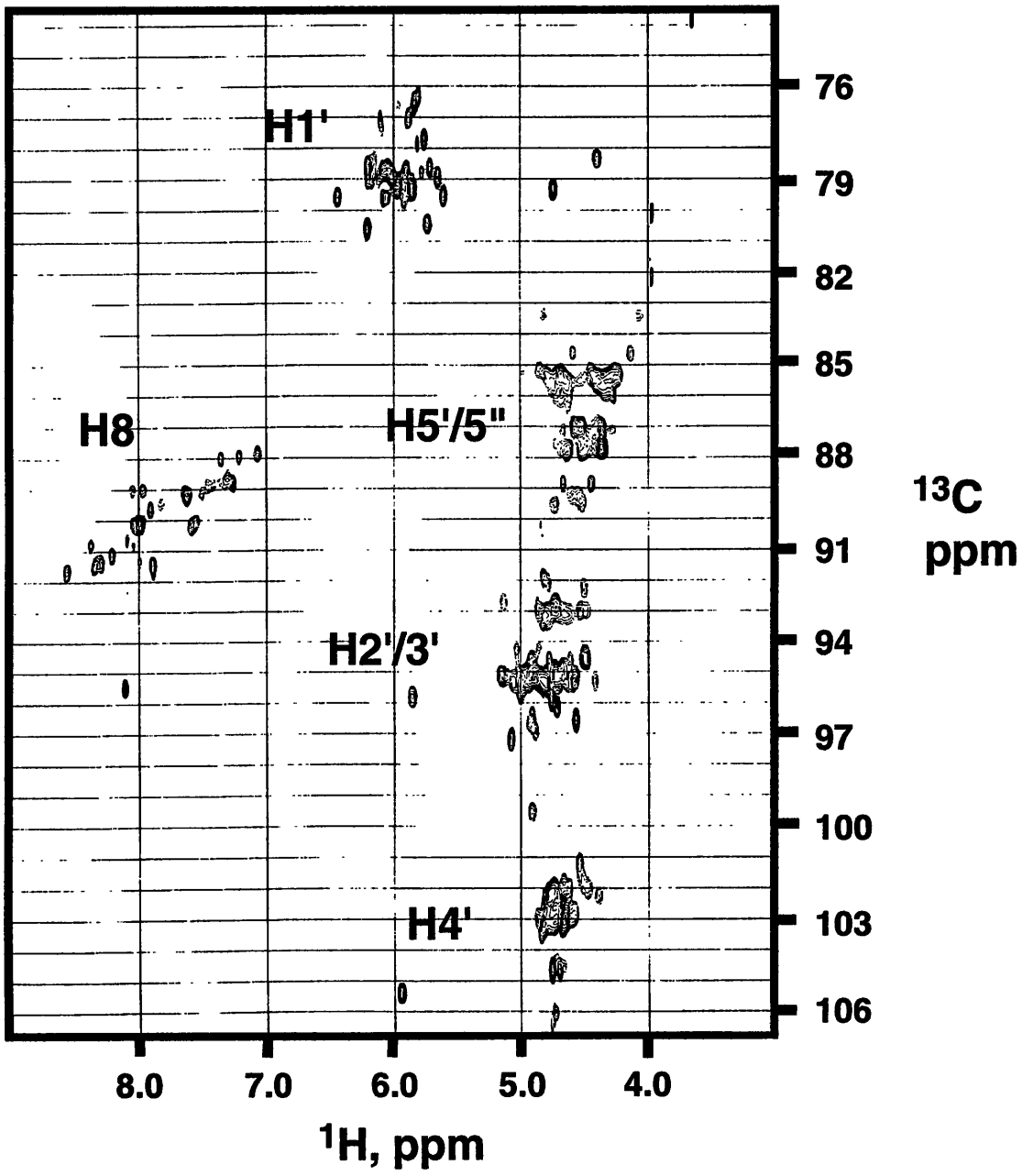


Figure 6.6: 2D-HSQC spectrum of the BS15-(^{13}C-G)Fr15 complex at 45 °C in 99.996% D2O. The guanosine nonexchangeable protons cluster into a distinct regions based upon the chemical shifts of that proton and the carbon to which it is directly attached. Each region should contain 24 peaks, one for each guanosine in Fr15 RNA. Note the high spectral dispersion in the H8 and H1' regions, and that the H2', H3', H4', and H5'/5" are resolved, unlike in homonuclear experiments.

likely due to the strong dipolar interaction between protons and the ^{13}C and ^{15}N heteronuclei to which they are directly bonded, which dramatically increases the rate of heteronuclear dipolar relaxation and, consequently, leads to broad linewidths (Wagner, 1993; Nietlispach *et al.*, 1996). Thus, the introduction of heteronuclei is the proverbial double-edged sword; while it dramatically improves the capability to maximize sequence-specific assignments, it can dramatically decrease the number of observable crosspeaks in large macromolecules and complexes via increases in relaxation rates.

Future directions for structural studies of the BS15-rRNA complex

One approach used to offset this effect in NMR spectroscopy of large proteins (>20 kD) is to use random fractional deuteration, resulting in narrower linewidths, reduced passive couplings, and suppression of spin diffusion (LeMaster, 1989; LeMaster, 1990; Sattler & Fesik, 1996). In NMR studies of the 37.5 kDa Trp repressor-DNA complex, the relaxation times of the C_α of the protein increased from 16.5 milliseconds in the fully protonated protein to 130 milliseconds in the 70% deuterated protein (Yamazaki *et al.*, 1994). Recently, the procedures for synthesis of $^{13}C/^{15}N$ labeled ribonucleotides have been modified in order to produce unlabeled and ^{13}C labeled nucleotides that are 50% randomly fractionally deuterated at all positions except the ribose H1' (the H1' position is deuterated at the 40% level due to nature of the metabolic pathways of *Methylophilus methylotrophus*) (Batey *et al.*, 1996). Currently, experiments are being undertaken to assess the improvement in spectral quality using BS15-Fr15(^{13}C-adenosine) samples that are undeuterated and with RNA that is 50% randomly fractionally deuterated. If random fractional deuteration provides significant improvements in the observed relaxation rates, as is observed for proteins (Wagner, 1993; Nietlispach *et al.*, 1996), then it might still be

possible to exploit some of the heteronuclear 3D-NMR experiments for the assignment phase of these studies.

It is clear from this work that high quality homonuclear spectra of the BS15-Fr15 complex can be obtained. Therefore, another approach to studying this complex may be to use homonuclear 2D-NMR spectroscopy in conjunction with many specific deuteration schemes that have been developed for RNA. One of the most useful regions of the RNA spectrum for making sequence specific assignments is the base aromatic to ribose anomeric region (shown in Figure 6.4b for this complex). The major problem associated with making assignments using this spectrum is the significant amount of peak crowding ("spectral crowding"). This problem can be relieved in a number of ways. A number of chemical techniques have been developed for the deuteration of the ribonucleotide bases at the purine H8, pyrimidine H6, and pyrimidine H5 positions to simplify this region of the NMR spectrum. The purine C8 position protons can be readily exchanged for deuterium by incubating rATP or rGTP in D_2O at 50 °C for 24 hours (Benevides *et al.*, 1984). The C5 position of pyrimidines can be exchanged for deuterium reacting rCTP or rUTP with deuterioammonium bisulfite under different conditions (Hayatsu, 1976) (the reaction of bisulfite with uracil does generate some amount of cytosine as a result of a deamination side reaction). Under more extreme conditions, it is possible that the protons on the C6 position of rUMP and rCMP can be exchanged for deuterons by incubating in DMSO-d_6/NaOD or MeOD (Rabi & Fox, 1973), and then phosphorylated using the techniques described in Appendix 2. Some of these techniques have been previously employed on studies of DNA (Brush *et al.*, 1988) and RNA (Puglisi *et al.*, 1990) oligonucleotides in order to make and confirm assignments by eliminating resonances in this region of the NMR spectrum. The regions of an NMR spectrum corresponding to the aromatic to ribose H2' and ribose H1' to ribose H2' can be similarly simplified using specific deuteration strategies that have recently been developed (Tolbert & Williamson, 1996). Another recently developed

strategy is the "NMR-window" concept in which RNA oligonucleotides are chemically synthesized using a combination of deuterated and protonated ribonucleotide phosphoramidites, yielding an RNA in which certain sequences are edited out of the NMR spectra by complete deuteration, which greatly simplifies the spectral overlap problem (Földesi *et al.*, 1992; Földesi *et al.*, 1996). The advantage to this technique is that one could observe only a single region of interest (for instance, only the phylogenetically conserved nucleotides in the three-way junction), but the fact that larger RNA oligonucleotides (>30 nucleotides) cannot yet be synthesized in large quantities easily remains a significant stumbling block to its routine application.

Of course, the alternative approach to determining the structure of the BS15-Fr15 complex by NMR spectroscopy is to use X-ray crystallography. Recent interest in RNA and RNA-protein complexes have led to the development on new techniques for their synthesis and crystallization. The work performed on developing and improving an NMR-amenable BS15-rRNA complex would certainly improve the chances of being able to grow diffraction-quality crystals of this complex, and efforts in the Williamson laboratory are underway to explore this avenue of structural studies.

NMR studies of a minimal three-way junction element

With the difficulties experienced in addressing the structure of Fr15 RNA in the presence of BS15, another model system was developed to study the folding of the critical three-way junction element. Based upon the studies presented in Chapter 5, an RNA was developed (Fr18, shown in Figure 6.7) that consists of the phylogenetically conserved three-way junction element which folds in the presence of divalent ions, but with most of the helix 22 deleted. To determine the effect of magnesium on the conformation of Fr18 RNA, a 1.5 mM sample in 10 mM sodium phosphate, pH 6.5, and 50 mM sodium chloride, was titrated at 25 °C with magnesium chloride in the NMR, while observing the

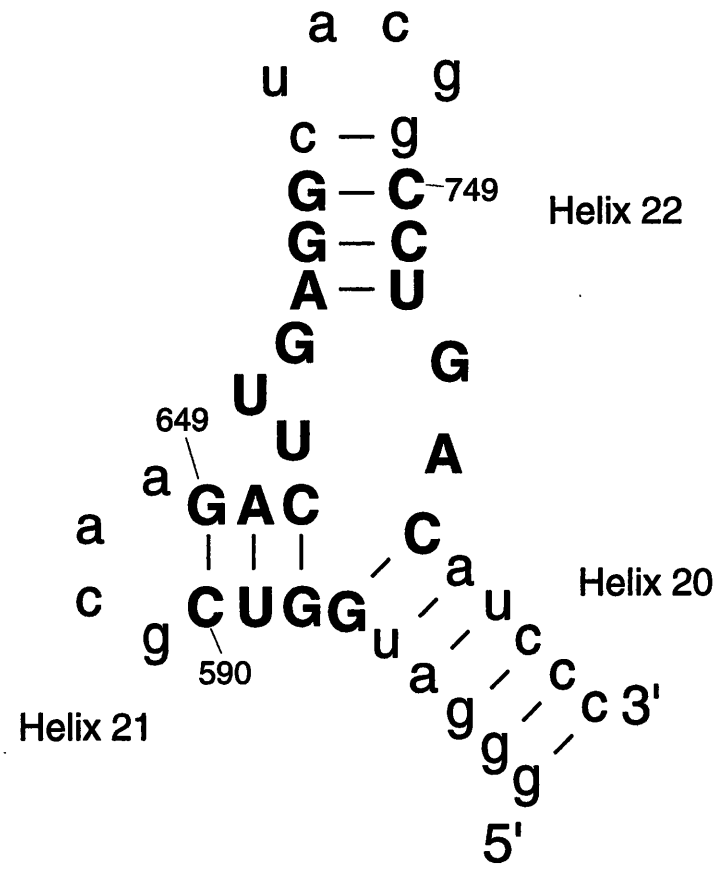


Figure 6.7: Fr18 RNA, element containing the phylogenetically conserved three-way junction element. Nucleotides which are present in 16S rRNA are represented as capital letters.

RNA imino spectrum (Figure 6.8). Upon addition of 15 mM magnesium chloride, the imino spectrum displays a number of new resonances that were absent without magnesium, indicative of a conformational change. Furthermore, the linewidths of the imino resonances are sharper in the presence of magnesium, suggesting that a more stable RNA structure has been formed in which the imino protons do not exchange as rapidly with bulk solvent. The conformational change upon addition of magnesium chloride to Fr18 RNA is also readily apparent in DQF-COSY spectra (Figure 6.9), taken at 35 °C. Most notably, in the magnesium bound form of Fr18, a number of new H5-H6 crosspeaks appear in the same positions in which they appear upon titration of Fr15 RNA with BS15 (these resonances are circled in Figure 6.9b). This strongly suggests that the conformation of the three-way junction element induced by the presence of BS15 or magnesium is very similar, since chemical shift is extremely sensitive to the chemical environment of the protons (Fr15 RNA at 45 °C displays similar behavior in the presence of magnesium—only a subset of H5-H6 resonances which shift in the presence of BS15 shift in the presence of magnesium). However, close inspection of the DQF-COSY spectrum of the magnesium bound form of Fr18 RNA indicates that this RNA in undergoing conformational exchange on the intermediate timescale. In the region of the spectrum shown in Figure 6.9, there should be 18 crosspeaks, one for each pyrimidine base. If an RNA is not undergoing any internal dynamic motions or conformational exchange, all of the H5-H6 crosspeaks should be observable and approximately the same intensity. Although there are clearly more crosspeaks in the DQF-COSY spectrum of Fr18 RNA in the presence of magnesium, the crosspeaks vary widely in their intensities and there are clearly not 18 peaks. Thus, the RNA is exchanging between multiple forms (probably between a "folded" and an "unfolded" form) on a timescale which causes extreme line broadening (see Figure 2.16), and thus makes this molecule an unacceptable model to study the three-way junction structure by NMR.

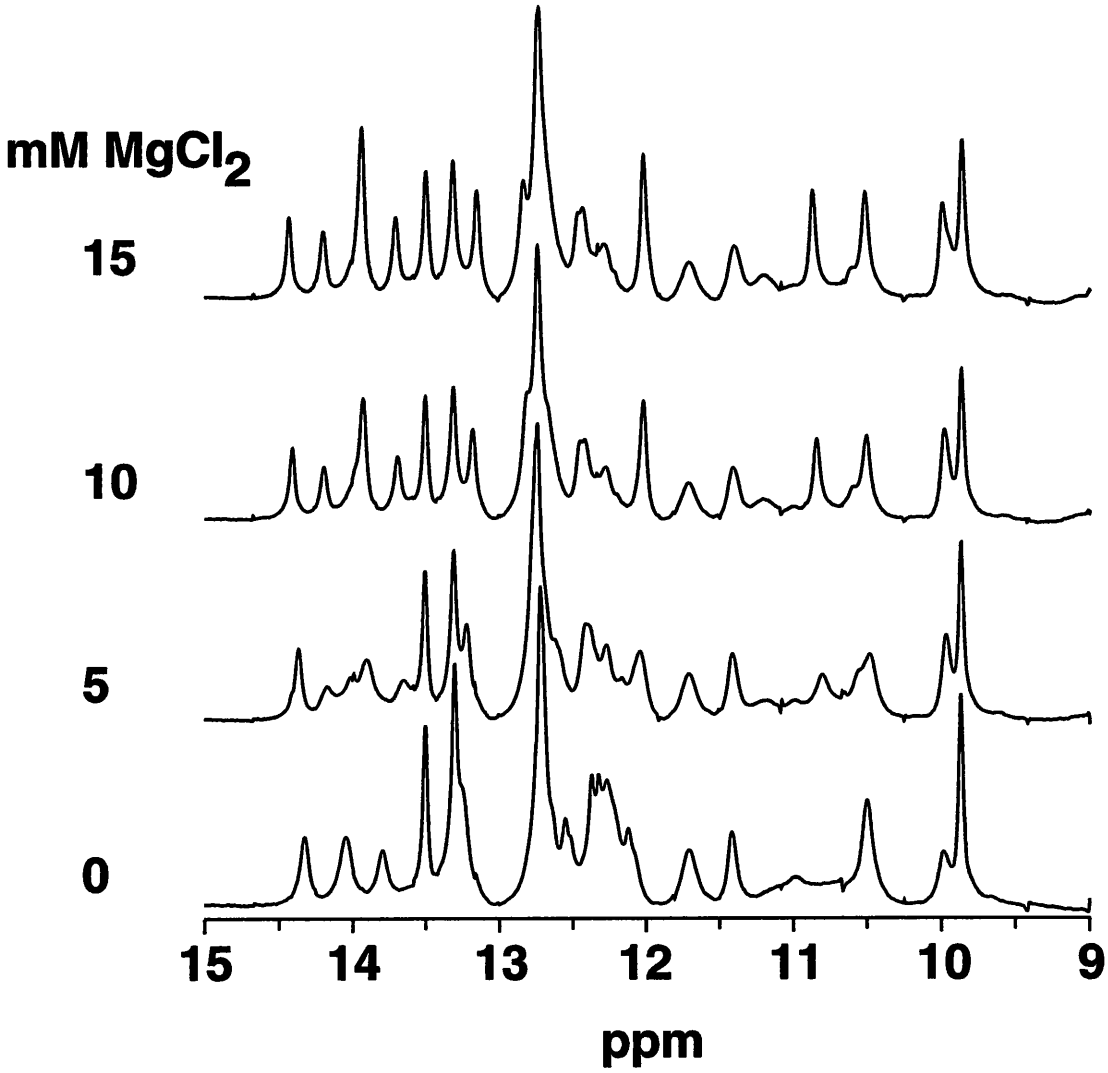


Figure 6.8: One-dimensional spectra of 1.5 mM Fr18 RNA in 90%/10% H_2O/D_2O showing a titration with magnesium chloride at 25 °C. A number of new imino peaks are evident in the RNA spectrum upon addition of 15 mM $MgCl_2$, evidence that this RNA undergoes a conformational change in the presence of divalent ions. Additional magnesium causes no further changes in the imino spectrum.

Figure 6.9 (next page): DQF-COSY spectra of Fr18 RNA acquired at 35 °C in (a) the absence and (b) the presence of magnesium chloride. The free RNA, which should display 18 pyrimidine H5-H6 crosspeaks, clearly shows only a small number of strong crosspeaks, indicating that the RNA is exhibiting severe conformational exchange. Although the situation is somewhat helped by the presence of magnesium, there are still too few crosspeaks in the lower spectrum. The circled resonances are H5-H6 crosspeaks that appear in the magnesium form of Fr18 RNA, which are also observed to appear when Fr15 RNA is titrated with BS15.

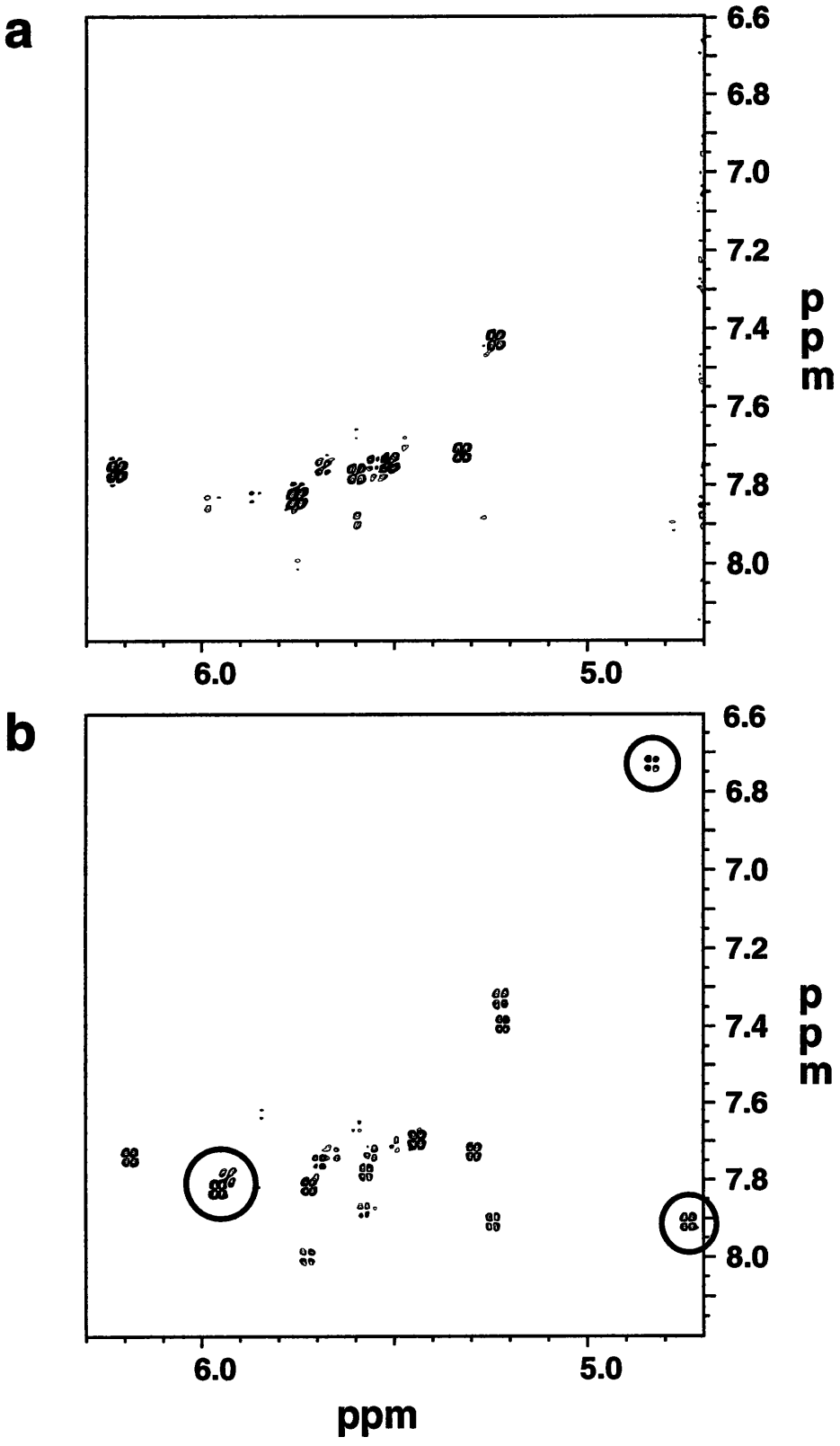

a
6.6
6.8
7.0
7.2
7.4
7.6
7.8
8.0
ppm
6.0
5.0
b
6.6
6.8
7.0
7.2
7.4
7.6
7.8
8.0
ppm
6.0
5.0
ppm

Figure 6.9

Chapter 7: Future directions

As detailed in the previous chapter, determination of the solution structure of the BS15-Fr15 complex is very much a work in progress, requiring many further experiments before an adequate model of the complex can be developed. While the solution structure is the major unanswered question posed by the preceding studies, it is by no means the only one worth pursuing in the future. Furthermore, the development of a biochemically well-defined S15-rRNA interaction serves as a logical springboard for detailed biochemical investigations of the RNA structure and protein-RNA interactions in the central domain of the *B. stearothermophilus* 30S ribosomal subunit.

Further RNA studies

Using a combination of mutagenesis, chemical modification interference analysis, and footprinting, the site of contact between S15 and 16S rRNA has been revealed, but the details as to which functional groups in 16S rRNA intimately participate in junction folding and protein-RNA recognition remain ambiguous. The fundamental problem lies in the types of chemicals that were utilized in the interference analysis. The RNA was probed with chemicals that either modify multiple functional groups (such as hydrazine, which alters the entire pyrimidine base), were unlikely to directly disrupt protein-RNA contacts of a minor groove-binding protein (like DEPC, which modifies at the purine N7 positions situated in the major groove), or introduce significant structural perturbations to the RNA in the locale of the modification site (such as CMCT, which contains two sterically bulky ring structures and DEPC, whose N7 modification cannot be accommodated in the major groove of a standard *A*-form RNA helix). Thus, it is extremely likely that most of the chemicals employed in these studies revealed the binding site of S15 via indirect effects. In

order to elucidate the specific elements of the RNA minor groove that are important for protein recognition, it will be essential to modify functional groups which lie on the minor groove side of the helix, such as the ribose 2' hydroxyl, the exocyclic amino group of guanosine, and the N3 position of purine bases. Since there are no chemical reagents that readily modify these positions, an alternative approach will be required.

In our early attempts to study the *E. coli* S15-16S rRNA interaction, a system was developed that would be ideal for addressing this issue. This system involves dividing the S15 binding site into three pieces, as shown in Figure 7.1. The three pieces of rRNA were synthesized using standard template-directed transcription using T7 RNA polymerase, and labeled at the 3' end. These three RNAs were mixed in equimolar quantities and annealed in a buffer containing 10 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, by heating to 90 °C for one minute and then slowly cooling to room temperature over the period of one hour. The three RNAs anneal to form a tripartite complex which contains the S15 binding site, as monitored using native gel electrophoresis (Figure 7.2). Based upon subsequent work with BS15, it is clear that this RNA complex should bind S15 with high affinity. Rather than transcribing all three strands with T7 RNA polymerase, one of the strands could be synthesized using a DNA synthesizer, allowing a single functional group within the RNA to be modified, such as the incorporation of inosine to remove a guanosine exocyclic amine or a deoxyribonucleotide to eliminate a ribose 2' hydroxyl group. Since the work in this thesis has clearly demonstrated the site of recognition for S15, only a limited set of singly modified RNAs needs to be synthesized in order to thoroughly probe the RNA folding and protein recognition. The ability of the singly substituted RNA complexes to fold or bind S15 could be easily tested using the same native gel electrophoresis techniques employed in this work. This approach has been successfully used by several groups to probe the HIV Rev-Rev response element interaction (Hamy *et al.*, 1993) and the interaction between *E. coli* alanine tRNA synthetase and ^{Ala}tRNA (Musier-Forsyth *et al.*, 1991; Musier-Forsyth & Schimmel, 1992). In each case, RNAs were synthesized with functional group

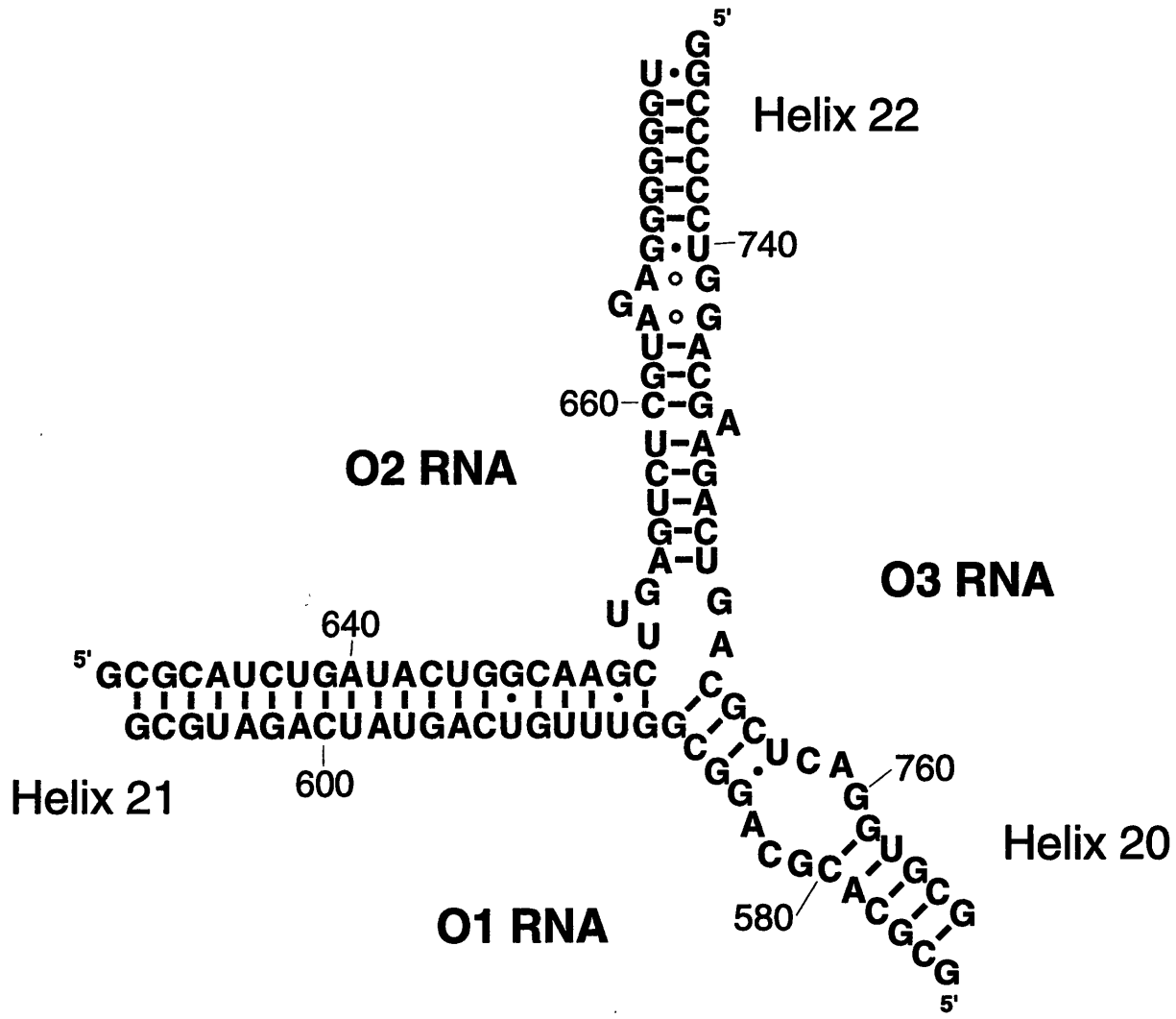


Figure 7.1: Three piece RNA complex comprising the *E. coli* S15 16S rRNA binding site.

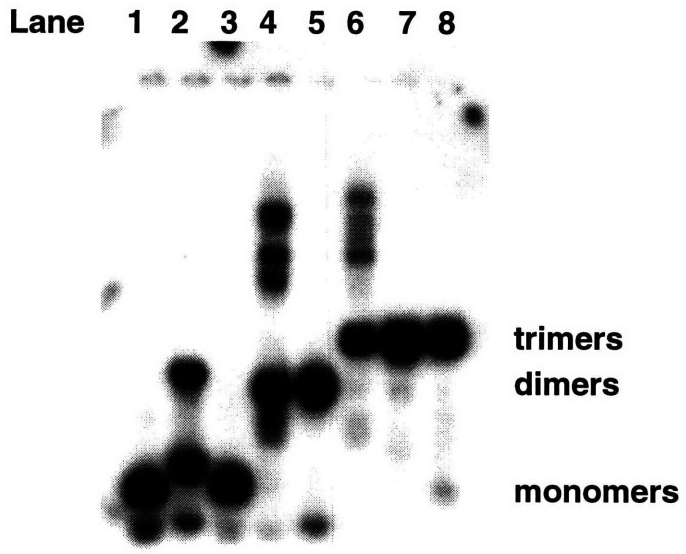


Figure 7.2: Formation of the tripartite S15 rRNA binding site as monitored using an 8% nondenaturing polyacrylamide gel. Reactions contain 50 nM unlabeled RNA (each strand) plus 50,000 cpm of the strand that has been 3' end-labeled in a buffer containing 10 mM Tris-Cl, pH 8.0, 10 mM $MgCl_2$ and annealed by heating to 80°C for 2 minutes and slowly cooling to 37°C. Lanes correspond to the reactions containing: 1, O1* RNA; 2, O2* RNA; 3, O3* RNA; 4, O1*/O2 RNA; 5, O1/O2* RNA; 6, O1*/O2/O3 RNA; 7, O1/O2*/O3 RNA; and 8, O1/O2/O3* RNA (asterick indicates the radiolabeled RNA fragment).

modifications that could not be accessed easily using other techniques, including the modifying agents commonly used in interference analysis, to generate detailed information about the contribution of specific RNA functional groups to the protein-RNA interaction.

Another powerful technique that has recently emerged for the analysis of protein-RNA interactions is *in vitro* genetic selection (SELEX). There have been a number of studies that have successfully utilized this technique to probe protein-DNA and protein-RNA interactions (Oliphant *et al.*, 1989; Turek & Gold, 1990; Bartel *et al.*, 1991). Briefly, this technique involves the synthesis of a pool of RNAs whose sequence has been randomized, selection of RNA variants that bind to a protein or ligand, and then PCR amplification of the selected nucleic acid pool (or their cDNA, as is the case for RNA). After multiple cycles of amplification and selection, a series of sequences emerge that are capable of binding the protein with high affinity. By sequencing a number of the RNA variants, an artificial phylogeny is created which provides constraints on the types of sequences that are capable of being recognized by the protein.

Although the extensive natural phylogeny of 16S rRNA has been very useful in the preceding studies, it suffers from two problems. First, within the S15 binding site, there are nucleotides which are conserved because they serve vital functions within the 30S subunit, but play little or no role in S15 recognition. A good example of this is G664, which is almost universally conserved in 16S rRNA phylogeny, but can be eliminated, along with the rest of the internal loop, without affecting S15 binding. Second, there may be covariation between the protein and the RNA. It has been determined that the A665-U751 base pair is extremely important for the S15-rRNA interaction, however, this base pair is often replaced by a G-C pair, particularly in archaeal 16S rRNAs. If this base pair participates in a critical protein-RNA contact, it is probable that there is a compensatory change in the corresponding S15 sequence. Thus, even though a base pair could be a critical protein-RNA contact, it could demonstrate variability in the natural phylogeny.

An artificial phylogeny would not suffer from these problems. During the *in vitro* selection experiment, the only selection pressure applied to the RNA is for the ability to bind to S15 with high affinity. Thus, conserved portions of this region of 16S rRNA that fulfill functional roles other than S15 binding would not be present in the resulting artificial phylogeny. Furthermore, since the protein does not have the opportunity to mutate during the SELEX experiment, covariation of amino acid-nucleotide contacts would not occur. This is not to say that SELEX does not suffer from problems of its own. The gel-mobility shift, which would certainly be the easiest enrichment technique for bound RNA, has been suggested to be prone to enrichment of sequences that aggregate (Bartel & Szostak, 1994). Also, it is known that S15 is capable of binding two different RNA species: an rRNA three-way junction and an mRNA pseudoknot. If the initial pool of nucleic acids contained a sixty nucleotide region that was completely randomized, the final pool may contain a mixture of pseudoknot RNAs and three-way junction RNAs, along with other potential classes of RNA motifs capable of folding into a structure recognized by S15, which would greatly complicate the analysis. On the other hand, the presence of other classes of binding competent RNAs would provide interesting insights into the variety of RNA motifs that could support specific S15 recognition. Thus, the design of the initial pool of RNA would be critically dependent upon what aspect of S15 binding is to be addressed in the experiment. Pools of nucleic acids with very limited complexity could easily be created to enrich for final sequences containing either the pseudoknot or three-way junction motif.

Another major question that these studies pose is the mechanism by which S15 recognizes rRNA; *i.e.* does RNA folding precede protein binding or vice versa? It has been demonstrated that the interaction of proteins with RNA promotes the formation and stabilization of RNA tertiary structure in functional ribozymes *in vivo* (Lambowitz & Perlman, 1990) via several different mechanisms. The cytochrome b pre-mRNA processing protein 2 (CBP2) interacts with the bI5 group I intron by recognizing weak, transiently formed tertiary structure (Weeks & Cech, 1995; Weeks & Cech, 1996),

whereas CYT-18, the *Neurospora crassa* mitochondrial tyrosyl-tRNA synthetase, actively promotes the formation of RNA tertiary structure by recognizing primary and secondary features of the ribosomal large subunit intron and promotes tertiary folding (Saldanha *et al.*, 1995; Caprara *et al.*, 1996). Thus, CBP2 and CYT-18 affect the kinetic pathways of folding in very different fashions. Similarly, formation of stable RNA tertiary structure and protein binding are intimately related in the S15-16S rRNA interaction. Understanding the mechanism by which S15 induces or stabilizes the three-way junction would provide insights into the processes by which ribosomal proteins help to fold the largely unstructured nascent 16S rRNA (Mandiyan *et al.*, 1989; Powers & Noller, 1995) into its active conformation in the functional 30S subunit.

Towards the solution structure of S15

During the course of this research, there has been an attempt to determine the solution structure of the uncomplexed *E. coli* and *B. stearothermophilus* variants of S15 using homonuclear and heteronuclear NMR spectroscopy. Even though high quality data were able to be collected for this protein, extensive sequence specific assignments were not able to be made (Parker *et al.*, unpublished observations). In brief, short stretches of sequence at the amino- and carboxyl-terminus of ES15 could be assigned, but the central forty amino acids were not assignable. The pattern of short-range NOEs observed for the assignable portion of the protein was consistent with those regions having significant alpha-helical character, but not a single long-range nOe could be identified that was indicative of a stable tertiary fold. Data collected on the *B. stearothermophilus* variant of S15 yielded similar results. The relatively unstructured nature of free S15 is clearly demonstrated in HMQC spectra acquired on ^{15}N-labeled BS15, shown in Figure 7.3a (Tao *et al.*, unpublished observations). Each peak in this spectrum corresponds to an amide proton in the protein backbone, asparagine side chain, or glutamine side chain. The poor

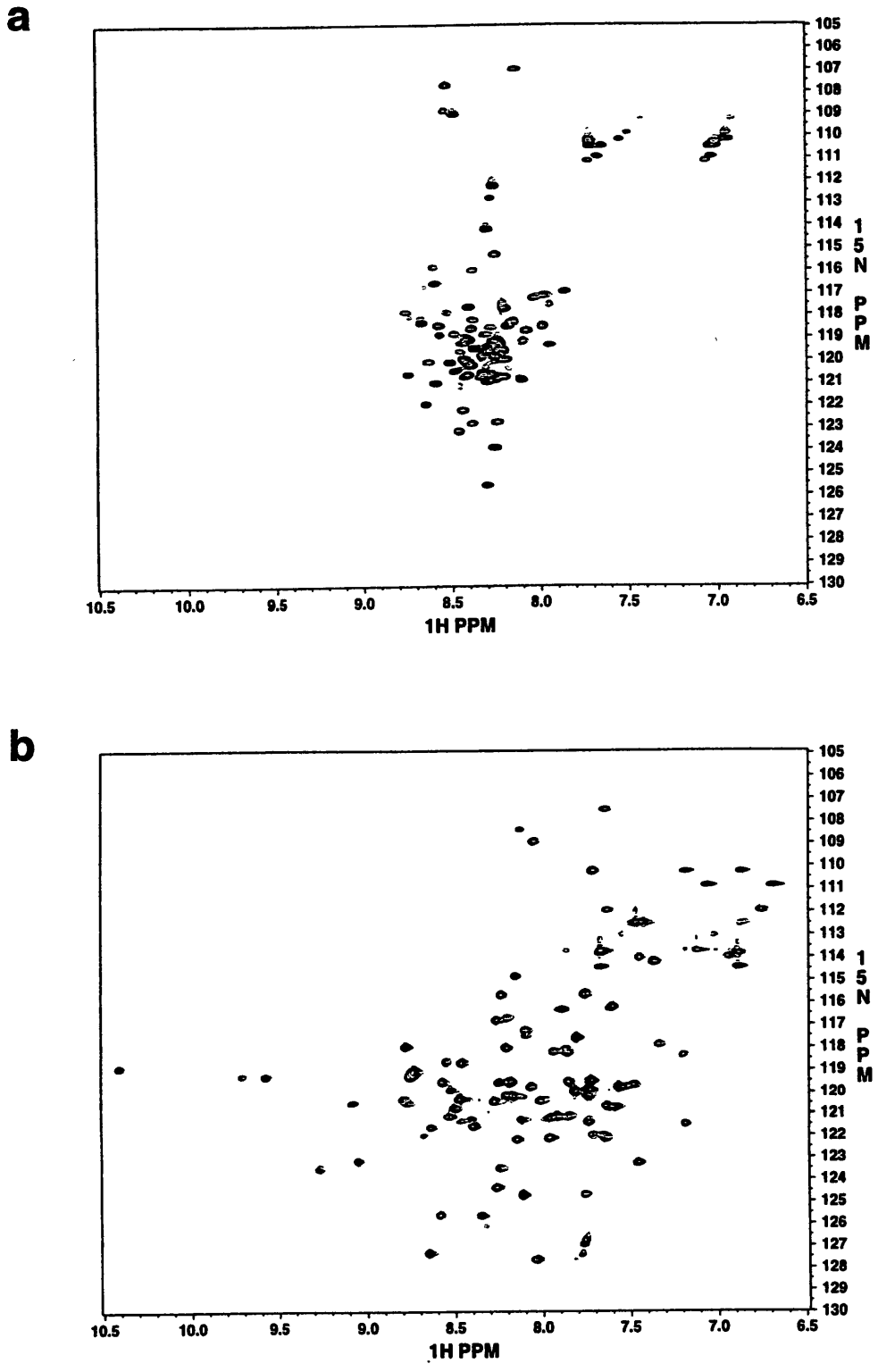


Figure 7.3: HMQC spectra of ^{15}N-labeled BS15 at 45 °C, (a) free and (b) complexed with a stoichiometric quantity of Fr15 RNA (Tao *et al.*, unpublished observations). Note the dramatic increase in the spectral dispersion of the amide resonances upon complexation with RNA.

spectral dispersion observed for these amides strongly suggests that this protein is poorly folded, and most likely does not have a unique tertiary structure. Upon complexation with an equimolar quantity of RNA containing its specific binding site (Figure 7.3b), the HMQC spectrum of the protein displays a high degree of spectral dispersion of the amide peaks, suggesting that the protein becomes folded upon binding RNA. This view of S15 structure is supported by protease protection experiments (Tao *et al.*, unpublished observations). In these studies, it was observed that BS15 is rapidly proteolyzed by endoprotease Glu-C after 30 minutes at 37 °C under non-denaturing conditions. However, in the presence of a stoichiometric amount of Fr15 RNA, the protein remains completely resistant to proteolysis for over 6 hours. The increased proteolytic stability of S15 in the presence of its specific RNA binding site provides additional evidence that this protein only achieves a stable tertiary fold upon binding RNA.

Currently, an effort is underway to determine which amino acids make contributions to RNA binding activity using an alanine scanning approach (Gibbs & Zoller, 1991). Originally, a single point mutant in BS15 was constructed by converting the universally phylogenetically conserved aspartic acid 47 to alanine (for sequence, see Figure 1.8) in order to test the specificity of the protein-RNA interaction (this amino acid was predicted to be in a loop region between alpha helices, which should contribute minimally to protein structure, but is known to be a site of protein-RNA contacts in other ribosomal proteins). This mutant protein was completely unable to bind to Fr1 RNA; at protein concentrations higher than 500 nM, the protein began forming lower mobility species indicative of nonspecific binding. Thus, elimination of a single carboxyl group in BS15 completely destroyed its ability to specifically interact with rRNA. Further mutagenesis studies are focused upon investigating the role of other conserved amino acids in S15 using the alanine scanning approach. Using a forthcoming solution structure of the unbound form of the *Thermus thermophilus* S15 variant (Berglund *et al.*, 1996), projection of the mutagenesis data upon the three-dimensional structure will provide important information

about which regions of S15 make important functional contributions to this protein-RNA interaction. This approach has been utilized successfully to determine the RNA binding face of a protein whose structure has been determined using X-ray crystallography (Predki *et al.*, 1995), but whose structure in the context of the protein-RNA complex is unknown.

Assembly of an RNP particle comprising the 30S central domain

As discussed in the introduction, the domain structure of the 16S rRNA manifests itself not only in the secondary structure of 16S rRNA, but also in the tertiary structure of the 30S subunits, as suggested by models developed independently by several groups. In these models the three major domains of the 16S rRNA are localized, along with their associated proteins, to distinct structural domains (see Figure 1.7). This has been supported by two lines of biochemical evidence. First, studies of the *in vitro* assembly pathway of the *E. coli* 30S subunit demonstrated that there are two independent nucleation events that occur during assembly (Nomura *et al.*, 1969), subsequently determined to be S4 and S7 (Nowotny & Nierhaus, 1988). Chemical footprinting studies have shown that the proteins whose binding to the assembling subunit is directly or indirectly potentiated by S4 and S7 interact almost exclusively with the same domain of rRNA, supporting the idea that the subdomains initially assemble autonomously. Second, partial ribonuclease digestion of 30S subunits yield a number of fragments from the three domains with a subset of ribosomal proteins still bound (Morgan & Brimacombe, 1972; Zimmerman, 1974). These rRNA fragments can be deproteinized and then reassembled with the same specific set of ribosomal proteins. More recently, rRNA fragments corresponding to the 5' and 3' domains of 16S rRNA have been synthesized by run-off transcription and reconstituted with the same ribosomal proteins that bind these domains in the 16S rRNA (Weitzmann *et al.*, 1993; Samaha *et al.*, 1994). These assemblages form discrete particles,

and at least in the case of the 3' domain RNP, show similar biological activity to that of the same domain in the context of the 30S subunit (Samaha *et al.*, 1994).

Since the ribosomal protein S15 strongly potentiates the assembly of the central domain (Held *et al.*, 1974), our understanding of this protein-RNA interaction serves as a natural stepping stone towards the understanding of a more complex RNP—the 30S central domain. Developing a central domain-like RNP would serve as an invaluable tool for detailed studies of the S6/S18-rRNA interaction, mechanisms of cooperativity between S15 and S6/S18, and crystallographic studies of this ribosomal substructure. Mutational and chemical footprinting studies of the central domain have demonstrated that a complex between S15, the S6/S18 heterodimer, and 16S rRNA can be formed (Gregory *et al.*, 1988; Svensson *et al.*, 1988). Furthermore, since S15 and the S6/S18 heterodimer contact adjacent regions of 16S rRNA, as demonstrated by Fe^{2+}-EDTA footprinting (Powers & Noller, 1995), and are close in space as seen in the neutron diffraction map of the 30S subunit (Capel *et al.*, 1987), this suggests a that these three proteins are intimately associated in the central domain. Using the same techniques employed in this work to generate a minimal S15-rRNA fragment, a minimal S15/S6/S18-rRNA quarternary complex could be readily developed as an ideal model system for probing the role of S15 in the assembly of the central domain of the 16S rRNA. Coupled with structural information, detailed studies of the assembly of the central domain could provide new insights into the assembly pathways of complex ribonucleoprotein particles.

Appendix 1: Synthesis and Purification of Biological Materials

Expression and purification of proteins

In this work, variants of the ribosomal protein S15 from *Bacillus stearothermophilus* and *Escherichia coli* have been utilized. Since these two proteins are highly homologous (58% amino acid identity between them, Figure 1.8), the purification protocols developed for each is very similar. It has been demonstrated that the 30S ribosomal subunit could be reconstituted from individual RNA and protein components purified under denaturing conditions, by isolating ribosomes from *E. coli*, separating the two ribosomal subunits, and removing the proteins from the 30S subunit using 8 M urea and 4 M LiCl (Traub *et al.*, 1971). This indicates that all of the ribosomal proteins can be denatured and refolded into their active form; in the reconstitution studies, this was accomplished by dialyzing away the urea and LiCl. More recently, efficient denaturing preparations of ribosomal proteins have been reported using ion exchange chromatography (Moore, 1979) and reverse phase HPLC (Kamp *et al.*, 1984; Cooperman *et al.*, 1988). The procedures developed to purify ES15 and BS15 in this thesis are derived from these protocols, with a few changes to optimize the purification.

All of these procedures for purification of ribosomal proteins exploited the fact that ribosomes, and thus all ribosomal proteins, are highly expressed in rapidly dividing cells. During log phase growth in *E. coli*, there are approximately 20,000 ribosomes per cell (Neidhardt, 1987). Thus, sufficient quantities of ribosomal proteins can be obtained through procedures that harvest purified ribosomes from bacterial cells and then purify the individual ribosomal proteins to homogeneity. Although this approach would have certainly yielded sufficient S15 for biochemical characterization, it would have been

difficult to purify the quantities necessary for NMR characterization. Furthermore, it would have been impractical to label the protein with ^{13}C and/or ^{15}N in order to apply heteronuclear NMR techniques to protein structure determination. Therefore, it was decided to construct overexpression vectors for the production of the *B. stearothermophilus* and *E. coli* variants of S15, as detailed below.

Bacillus stearothermophilus protein S15

Construction of the BS15 overexpression plasmid

The gene sequence of the *Bacillus stearothermophilus* ribosomal protein S15 (BS15) is not known. Instead, the protein sequence of BS15, along with many other of the *B. stearothermophilus* ribosomal proteins, was obtained by direct sequencing of purified ribosomal proteins (Arndt *et al.*, 1991). Based upon this information, a synthetic gene for protein expression was made using synthetic oligonucleotides.

An overexpression plasmid containing a gene encoding the BS15 protein was constructed based upon the reported amino acid sequence [EMBL #P05766] (Arndt *et al.*, 1991). The gene was constructed from four synthetic oligonucleotides, two sense and two antisense strands, with a fifteen base overlap between the upstream and downstream neighbors. Oligonucleotides were synthesized on an Applied Biosystems PCR Mate 391 DNA synthesizer, using standard phosphoramidite chemistry. Synthetic DNA oligonucleotides were cleaved from the support resin and deprotected by incubating the solid support in 1.5 ml aqueous ammonium hydroxide for 12-16 hours at 65 °C in sealed glass vials. The aqueous solution was lyophilized to dryness to remove the ammonia prior to purification of the DNA oligonucleotides on a 12% denaturing polyacrylamide gel.

The strategy employed for the construction of the BS15 overexpression vector is shown in Figure A1.1. The sequence contained *Eco*RI and *Hind*III restriction sites

engineered at the 5' and 3' ends for cloning, and the codons were selected upon the basis of the *E. coli* preferred codon usage (de Boer & Kastelein, 1986; Williams *et al.*, 1988) (Figure A1.2) . In separate reactions, the oligonucleotides P1, $^{5'}$GGCGGGGAATTC-ATGGCTCTGACCCAGGAACGCAAACGCGAAATCATCGAACAGTTCAAAGTTCA-CGAAAACGACACCGGTTCCCCGG, and P2, $^{5'}$GCGAACGCGCAGGTGTTCGTT-CAGGTTGTTGATCTGTTCGGTCAGGATAGCGATCTGAACTTCCGGGGAACCGG-TG, and the oligonucleotides P3, $^{5'}$CACCTGCGCGTTCGCAAAAAAGACCACCAC-TCCCGCCGCGGTCTGCTGAAAATGGTTGGTAAACGCCGCCGCCTG, and P4, $^{5'}$GCCCTCAAGCTTTCACTAGCGGCGCAGACCCAGTTTTTCAACGATTTCGCGG-TAGCGAGCAACGTCTTTGTT-GCGCAGGTAAGCCAGCAGGCGGCGGCG, were annealed and extended using 750 ng of each strand and AmpliTaq DNA polymerase (Perkin Elmer-Cetus) with 20 cycles of denaturing at 94 °C for 1 minute, annealing at 55 °C for 1 minute, and extension at 72 °C for 2 minutes (Innis *et al.*, 1990). The products of each reaction were purified on 8% nondenaturing 29:1 acrylamide:bisacrylamide gels in TBE buffer (0.09 M Tris-borate, 2 mM EDTA, pH 8.1), the correct product excised from the gel and soaked into TE buffer (10 mM Tris-Cl, 1 mM EDTA, pH 8.1) overnight. An amplification reaction using the P1 and P4 fragments as the 5' and 3' primers in a PCR reaction using the above cycle and 1 µl of the eluted DNA from the previous reactions was used to construct the intact gene. For cloning, the PCR product was desalted using a 30,000 molecular weight cut- off microconcentrator and cleaved with *Eco*RI and *Hind*III restriction enzymes in a reaction containing 10 mM Tris-Cl, pH 7.9, 10 mM $MgCl_2$, 50 mM NaCl, 1 mM DTT at 37 °C. The insert was purified on a 2% LMP agarose gel and extracted from the gel using a QIAEX kit (Qiagen). The product was ligated into pKK223-3 (Pharmacia) linearized with *Eco*RI and *Hind*III using T4 DNA ligase at 16 °C for 16 hours. pKK223-3 is an vector carrying a multiple cloning site immediately downstream of the strong, regulateable *tac* promoter for protein expression (Amann *et al.*, 1983) and the rrnB transcription terminators (Brosius & Holy, 1984). The ligation reaction was used to

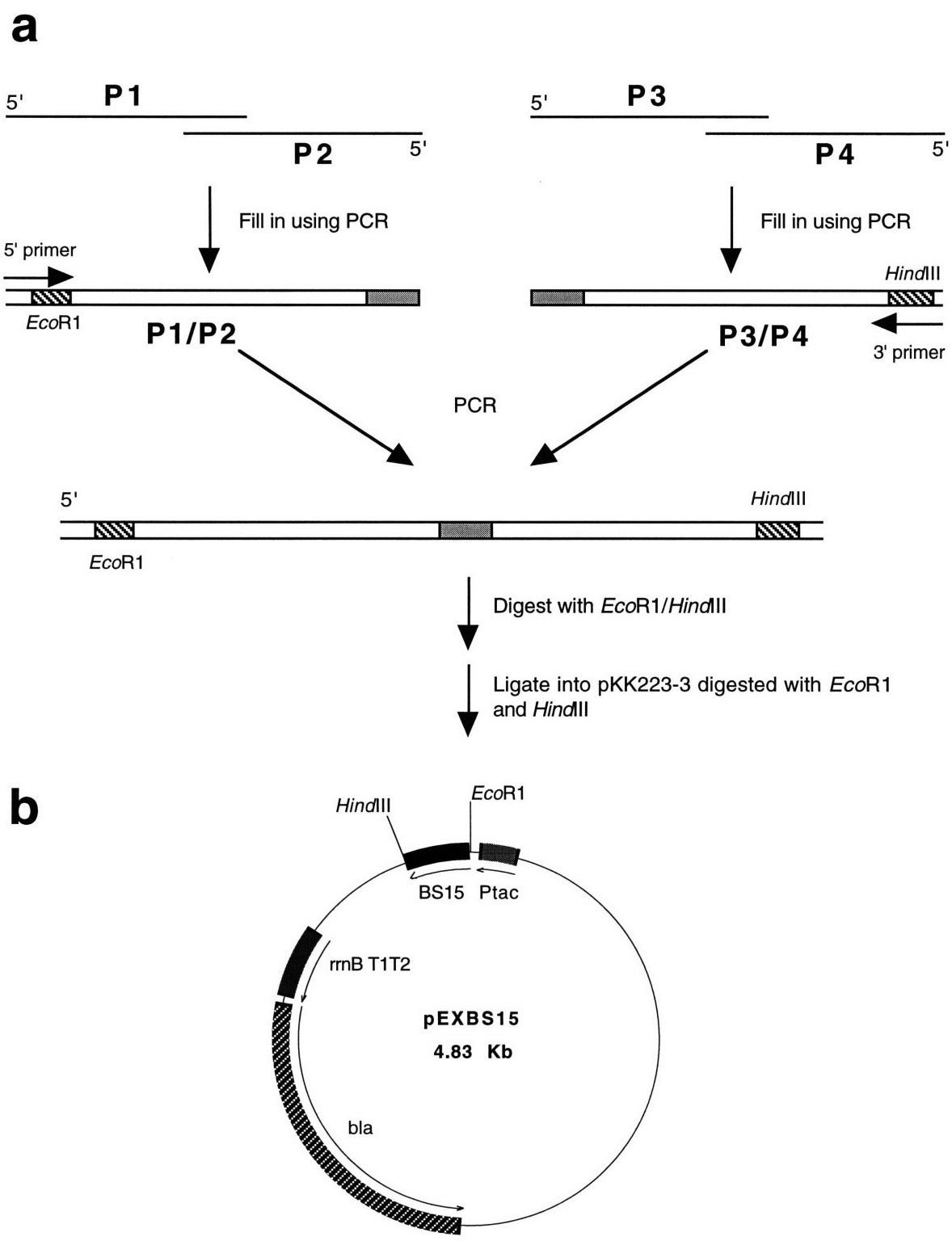


Figure A1.1: (a) Construction of the BS15 gene using synthetic oligonucleotides. The *Eco*R1 and *Hind*III sites at the 5' and 3' ends of the gene are used for cloning into pKK223-3. (b) Map of the BS15 expression vector, pEXBS15.

```
ATG GCT CTG ACC CAG GAA CGC AAA CGC GAA ATC ATC GAA
Met Ala Leu Thr Gln Glu Arg Lys Arg Glu Ile Ile Glu

CAG TTC AAA GTT CAC GAA AAC GAC ACC GGT TCC CCG GAA
Gln Phe Lys Val His Glu Asn Asp Thr Gly Ser Pro Glu

GTT CAG ATC GCT ATC CTG ACC GAA CAG ATC AAC AAC CTG
Val Gln Ile Ala Ile Leu Thr Glu Gln Ile Asn Asn Leu

AAC GAA CAC CTG CGC GTT CGC AAA AAA GAC CAC CAC TCC
Asn Glu His Leu Arg Val Arg Lys Lys Asp His His Ser

CGC CGC GGT CTG CTG AAA ATG GTT GGT AAA CGC CGC CGC
Arg Arg Glu Leu Leu Lys Met Val Gly Lys Arg Arg Arg

CTG CTG GCT TAC CTG CGC AAC AAA GAC GTT GCT CGC TAC
Leu Leu Ala Tyr Leu Arg Asn Lys Asp Val Ala Arg Tyr

CGC GAA ATC GTT GAA AAA CTG GGT CTG CGC CGC TAG TGA
Arg Glu Ile Val Glu Lys Leu Gly Leu Arg Arg AMB OPA
```

Figure A1.2: Sequence and translation of the constructed BS15 gene. The initiator ATG codon is immediately downstream of the *Eco*R1 site in pEXBS15, and the opal stop codon is immediately downstream of the *Hind*III site (see figure A1.1). The codon used in the construction of this gene are those that are most frequently utilized in the *E. coli* genome.

transform into JM109 *E. coli* cells using the Hanahan chemical transformation protocol (Hanahan, 1983). Transformants were selected on the basis of ampicillin resistance, and the plasmids were checked for the presence of an insert by restriction analysis. Using a sequencing primer made for pKK223-3, $^{5'}$GACAATTAATCATCGGCTCG, plasmids containing the insert were sequenced by the dideoxynucleotide sequencing method using Sequenase 2.0 (United States Biochemical). Clones containing the correct gene sequence were tested for expression of a protein corresponding to approximately 10 kDa. The selected clone, pEXBS15/JM109, was used for all subsequent expression and purification of *B. stearothermophilus* S15 protein.

Purification of the BS15 Protein

pEXBS15/JM109 was grown in 2 liters of Terrific Broth (Sambrook *et al.*, 1989) containing 50 μg/ml of ampicillin at 37 °C. The culture was induced with 1 mM IPTG when the culture reached an O.D. at 660 nm of 0.5 and allowed to grow for another 12 hours. The cells were centrifuged in a Beckman JA-10 rotor for 15 minutes at 8000 rpm. The cell pellets, approximately 10 grams, were resuspended in Sonication buffer (6 M urea, 50 mM NH_4OAc, pH 8.1, 5 mM DTT, and 1 mM EDTA) and sonicated with 1 minute bursts for a total of 30 minutes while on ice. The pH of the sonicate was reduced to 5.6 with glacial acetic acid and spun in a JA-10 rotor for 30 minutes at 8000 rpm. The supernatant was applied to a 50 ml Toyopearl SP column (Supelco) pre-equilibrated with Column buffer (6 M urea, 30 mM NH_4OAc, pH 5.6) at 4 °C with a flow rate of 1.5 ml/min. After application, the column was washed with Column buffer until the absorbance at 280 nm returned to baseline. The protein was eluted with a 600 ml linear gradient of 0 to 0.5 M NaCl in Column buffer. The fractions were analyzed by SDS-PAGE for the presence of BS15 and those fractions containing the protein were pooled and dialyzed overnight against 2% acetic acid at 4 °C. The protein solution was lyophilized and

resuspended in 5 ml Column buffer. One ml aliquots were applied to a C18 HPLC column (2.2x25 cm, Vydac) equilibrated in ddH_2O/0.1% trifluoroacetic acid (TFA) with a flow rate of 5 ml/min and eluted with a linear gradient of 0-50% acetonitrile/0.1% TFA over 120 minutes. The peak containing BS15 was collected and lyophilized.

The lyophilized protein was resuspended in 10 mM MOPS buffer, pH 7.0, and exhaustively dialyzed against two liters of this buffer overnight in the cold room. This protein solution was applied to a Progel-TSK CM3SW HPLC column (Supelco) equilibrated in 10 mM MOPS buffer, pH 7.0, with a flow rate of 1.0 ml/min. This column acted both as a weak cation exchange column and as a size exclusion column, since the column matrix is derivatized on the inside of beads with a 30 kDa exclusion limit. The protein was eluted from the column using a linear gradient of 0 to 1 M KCl in 10 mM MOPS buffer, pH 7.0, over 120 minutes. The fractions containing BS15 were collected, pooled, and dialyzed overnight against 2% acetic acid at 4 °C. This protein was then subjected to one more round of purification using the preparative C18 HPLC column, as described above.

The BS15 protein purified in this fashion was determined to be at least 99% pure by a Coomasie-stained SDS-PAGE gel (Figure A1.3). The identity of the resulting protein was verified by N-terminal sequencing of the first 6 amino acids and by amino acid composition analysis. The extinction coefficient used for calculation of the protein concentration was 46,900 M^{-1} cm^{-1} at 230 nm. The protein, under native conditions, tended to slowly aggregate, and therefore, the protein was stored long-term as a lyophilized powder.

Prior to an experiment, a concentrated stock of the protein was made in a solution containing 4 M urea, 10 mM K-HEPES, pH 7.5, 50 mM potassium acetate, 0.1 mM EDTA, and 50 μg/ml heparin; short term storage of the protein in the presence of urea and heparin ameliorated the aggregation problem. This stock was stored at 4 °C for no more than a week before use in binding experiments; protein stored for a longer period than this

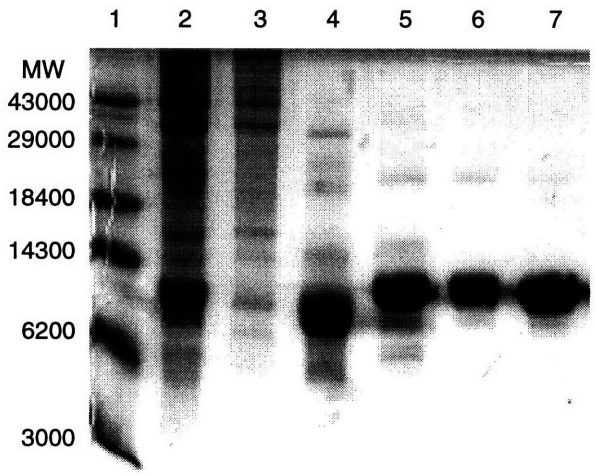


Figure A1.3: Coomasie-stained 17.5% SDS-polyacrylamide gel of BS15 protein at various stages in the purification procedure. The lanes correspond to (1) low molecular weight standards, (2) urea-lysed JM109/pEXBS15 after 4 hours of induction with IPTG, (3) SP column flow through, (4) SP elution fraction containing BS15, (5) RP-HPLC elution fraction containing BS15, (6) CM-HPLC elution fraction containing BS15, (7) final purified BS15 after second RP-HPLC purification. Numbers to the left indicate molecular weight in daltons.

gave reduced binding affinities and markedly sharper binding transitions from free to bound RNA.

Escherichia coli ribosomal protein S15

PCR amplification of the protein from E. coli genomic DNA

Unlike BS15, the gene sequence for the *E. coli* ribosomal S15 protein (rpsO) was known, and therefore, the gene could be amplified directly from genomic DNA. The genomic DNA was prepared from the *E. coli* strain MRE 600 by pelleting 1.5 ml of a 5 ml overnight culture, resuspended the cells in 100 µl ddH_2O, and boiling for five minutes in a water bath. This solution was then cooled on ice for 2 minutes, centrifuged, and the supernatant containing genomic DNA transferred to a fresh tube (Clackson *et al.*, 1991). Oligonucleotides used for PCR amplification were synthesized on an Applied Biosystems PCR Mate 391 DNA synthesizer, using standard phosphoramidite chemistry. The 5' primer, $^{5'}$TCAGAGGAATTCATGTCTCTAAGTACTGAAGCAACAGCT, had an *Eco*RI restriction site attached immediately upstream of the ATG initiator codon, and the 3' primer, $^{5'}$CCTAAGAAGCTTAATTTAGCGACGCAGACCCAGGCGCTCGAT, had an additional stop codon attached immediately downstream of the natural stop codon along with a *Hind*III restriction site. To amplify the S15 gene, a reaction containing 1 µg of each primer, 1 µl of MRE 600 genomic DNA were added to PCR buffer (10 mM Tris-HCl, pH 8.8, 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% (w/v) gelatin) and denatured for five minutes at 96 °C. After the initial denaturation of the genomic DNA, one unit of AmpliTaq DNA polymerase was added, and the reaction was amplified using 30 cycles of denaturing at 94 °C for 1 minute, annealing at 60 °C for two minutes, and extension at 72 °C for two minutes (Innis *et al.*, 1990). A small quantity of this reaction was electrophoresed

alongside molecular weight markers on a 1% agarose gel, and a single band was observed corresponding to a DNA fragment of the expected length of 294 base pairs.

The construction of the ES15 expression plasmid, pEXES15, was performed using the same protocol as described above for the construction of pEXBS15, and subsequently transformed into the *E. coli* strain JM109.

Expression controls

Prior to large scale production of ES15 and BS15 for biochemical and NMR studies, the expression conditions were optimized for maximal protein production. First, expression was tested in both rich medium (Luria Broth and Terrific Broth, Sambrook *et al.*, 1989) and minimal medium (glycerol-minimal salts medium supplemented with 5 μg/ml thiamine, since JM109 is deficient in the ability to synthesize this vitamin). Protein expression in minimal medium is particularly important for isotopic labeling of the protein with ^{13}C and/or ^{15}N for used in structural studies by NMR. In all media tested, a strong band appeared at the correct molecular weight in a Coomasie stained SDS-polyacrylamide gel upon induction of protein expression with IPTG, indicative of high levels of protein expression.

The second variable tested was the time of induction. IPTG induction of S15 expression in exponentially growing bacterial cultures typically caused the growth to slow down considerably, resulting in overall lower cell densities. However, if induction occurred close to saturation of the culture, very little protein expression was observed. Thus, the time of addition of IPTG was optimized to ensure maximal protein expression along with maximal cell density at the end of the growth. For expression of S15 in Terrific Broth, it was determined the optimal induction time occurred when the culture reached an O.D. at 660 nm of 0.5, and for expression in glycerol-minimal salts medium, the optimal time of induction occurs at 1.0 O.D.

Finally, protein expression levels as a function of time after the induction with IPTG was monitored. In Terrific Broth and glycerol-minimal salts medium, maximal protein expression was achieved 2-3 hours after induction, as observed using a Coomasie stained SDS-polyacrylamide gel (Figure A1.4). Since the level of protein expression did not diminish upon further growth (4-8 hours), the cell cultures were typically harvested eight hours after induction in order to maximize the cell density.

Purification

The procedure used for the production of *E. coli* S15 is similar to the procedure used for the production of BS15, except for a few minor changes. The initial column used for purification of ES15 was a 50 ml Toyopearl CM column (Supelco) using the same column buffer, and eluting with a 1 liter gradient of 0 to 0.5 M NaCl in Column Buffer. The fractions containing ES15, as determined by SDS-PAGE, were collected and purified using a preparative C18 HPLC column as described previously. After this step, the protein appeared to be >99% pure by SDS-PAGE, which was sufficient for the biochemical experiments in which this protein was used.

T7 RNA Polymerase

For these studies, we required milligram quantities of T7 RNA polymerase. Therefore, it was necessary to express and purify this enzyme from an overexpression strain. The T7 RNA polymerase was purified from the overproducing strain pAR1219/BL21 provided by F. W. Studier (Davanloo *et al.*, 1984; Grodberg & Dunn, 1988). Purification of this enzyme was accomplished using exactly the protocol described elsewhere (Grodberg & Dunn, 1988).

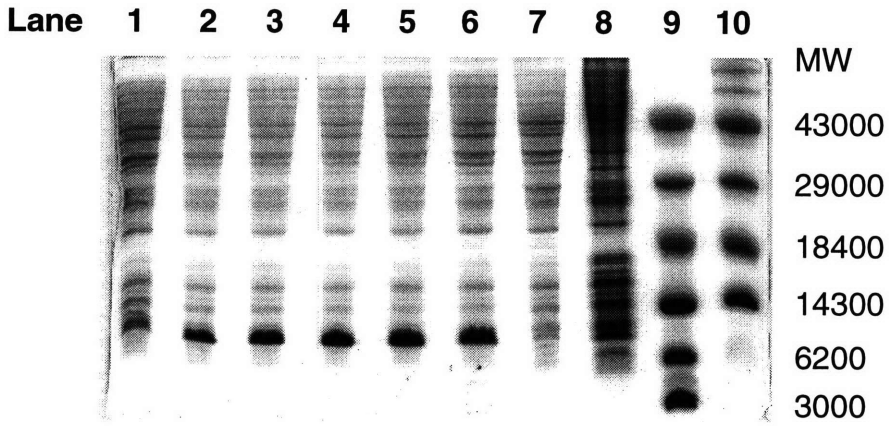


Figure A1.4: Coomasie-stained 15% SDS-polyacrylamide gel of a timecourse of ES15 expression. For each timepoint, 2 O.D. of culture was harvested, pelleted, and resuspended in 100 µl of SDS-PAGE load buffer and boiled. For each timepoint, 4 µl of the lysed cells was loaded into a well, such that for each timepoint an equal amount of cellular protein was electrophoresed. The lanes correspond to (1) JM109/pEXES15 at 0 hours after induction with IPTG, (2) one hour after induction, (3) 2 hours after induction, (4) 4 hours after induction, (5) 6 hours after induction, (6) 8 hours after induction, (7) JM109 cells (without the ES15 overexpression plasmid), (8) *E. coli* 70S ribosomal proteins, (9) BRL low molecular weight markers, (10) BRL high molecular weight markers. Protein expression reaches a maximum after four hours of induction with IPTG.

Synthesis and Purification of RNAs

Purification of Bacillus stearothermophilus 16S rRNA

Bacillus stearothermophilus (ATCC 7593) was grown in 500 ml Luria Broth (Sambrook *et al.*, 1989) until late log phase growth (O.D. 660 of 1.0) at 60 °C, and harvested by centrifugation in a Beckman JA-10 rotor for 15 minutes at 8000 rpm. Total cellular nucleic acids were harvested using a procedure for the isolation of RNA from gram-positive bacteria (Reddy & Gilman, 1989). The cells were resuspended in 25 ml of Lysis Buffer (30 mM Tris-Cl, pH 7.4, 100 mM NaCl, 5 mM EDTA, 1% (w/v) SDS, 100 µg/ml proteinase K), frozen on dry ice and thawed. The thawed cells were sonicated three times with 10 second bursts, and the lysate incubated for one hour on ice. Cellular proteins were removed by extracting the lysate three times with an equal volume of 25:24:1 phenol:chloroform:isoamyl alcohol and removing the aqueous layer after each extraction, followed by one extraction against an equal volume of 24:1 chloroform:isoamyl alcohol. The aqueous phase was precipitated by adding 0.1 volume 3 M sodium acetate, pH 5.2, and 2.5 volumes of 100% ethanol. The total cellular nucleic acids were resuspended in 1 ml TE buffer. 250 µl of this solution was electrophoresed in a 1% agarose gel in TAE buffer in the presence of ethidium bromide. The band corresponding to 16S rRNA was excised and eluted from the gel using an Elutrap (S&S). The RNA was ethanol precipitated, extracted twice against an equal volume of 25:24:1 phenol:chloroform:isoamyl alcohol and reprecipitated with ethanol. The agarose gel purification step was repeated once more to improve the purity of the 16S rRNA. To calculate the concentration of 16S rRNA, an extinction coefficient of 1.3×10^7 M^{-1} cm^{-1} at 260 nm was used (Schwarzbauer & Craven, 1981).

Construction of plasmids containing RNA genes

A fragment of 16S rRNA corresponding to nucleotides 585-756 was amplified from *B. stearothermophilus* (ATCC 7593) genomic DNA using a primer complementary to the 5' end, $^{5'}$GGCGAATTCGAGCTCGCGGTCCCTTAAGTCTGATGTGAAAGC, containing an *Eco*RI site, and an antisense primer complementary to the 3' end of the gene, $^{5'}$GCCGCCAAGCTTCCCGGGCGAATTCAAACTCGCGTCAGGTGCAGGCCAGAGA GCCGCC, containing a *Sma*I site for linearization of the plasmid for run-off transcription and a *Hind*III site for cloning. The gene fragment was amplified in a 100 µl PCR reaction containing 1 µM of each primer, 0.2 mM of each of the deoxynucleotide triphosphates, 10 mM Tris-Cl, pH 8.8, 50 mM KCl, 1.5 mM $MgCl_2$, 0.1% (w/v) gelatin, and 2.5 U of AmpliTaq DNA polymerase. Amplification was carried out using 30 repetitions of the following cycle: denaturation for 1 minute at 94 °C, annealing for 1 minute at 50 °C, and extension for 2 minutes at 72 °C. For cloning, the PCR product was desalted using a 30,000 MW cut-off microconcentrator and cleaved with *Eco*RI and *Hind*III restriction enzymes (New England Biolabs). The insert was purified on a 2% LMP agarose gel and purified using a QIAEX kit (Qiagen). The PCR product was ligated into pGEM 3zf(-) (Promega) cut with *Eco*RI and *Hind*III using T4 DNA ligase and incubated overnight at 16 °C. The ligation reaction was transformed into JM109 according to the standard Hanahan chemical transformation protocol (Hanahan, 1983). Transformants were screened on the basis of ampicillin resistance and by restriction analysis. The sequence of pB16SFr1 containing the insert was checked by sequencing using the universal sequencing primer and the dideoxynucleotide sequencing method (the structure of pB16SFr1 and other pGEM 3zf(-) derived plasmids are illustrated in Figure A1.5a). The plasmids containing the sequences for Fr2 and Fr3 were constructed in the same fashion.

To generate internal deletions, a recombinant PCR approach was utilized (Innis *et al.*, 1990). The upstream segment of the Fr1 gene was amplified using a sense primer complementary to the 5' end of the Fr1 gene, $^{5'}$GGCGAATTCGAGCTCGCGGTCCC,

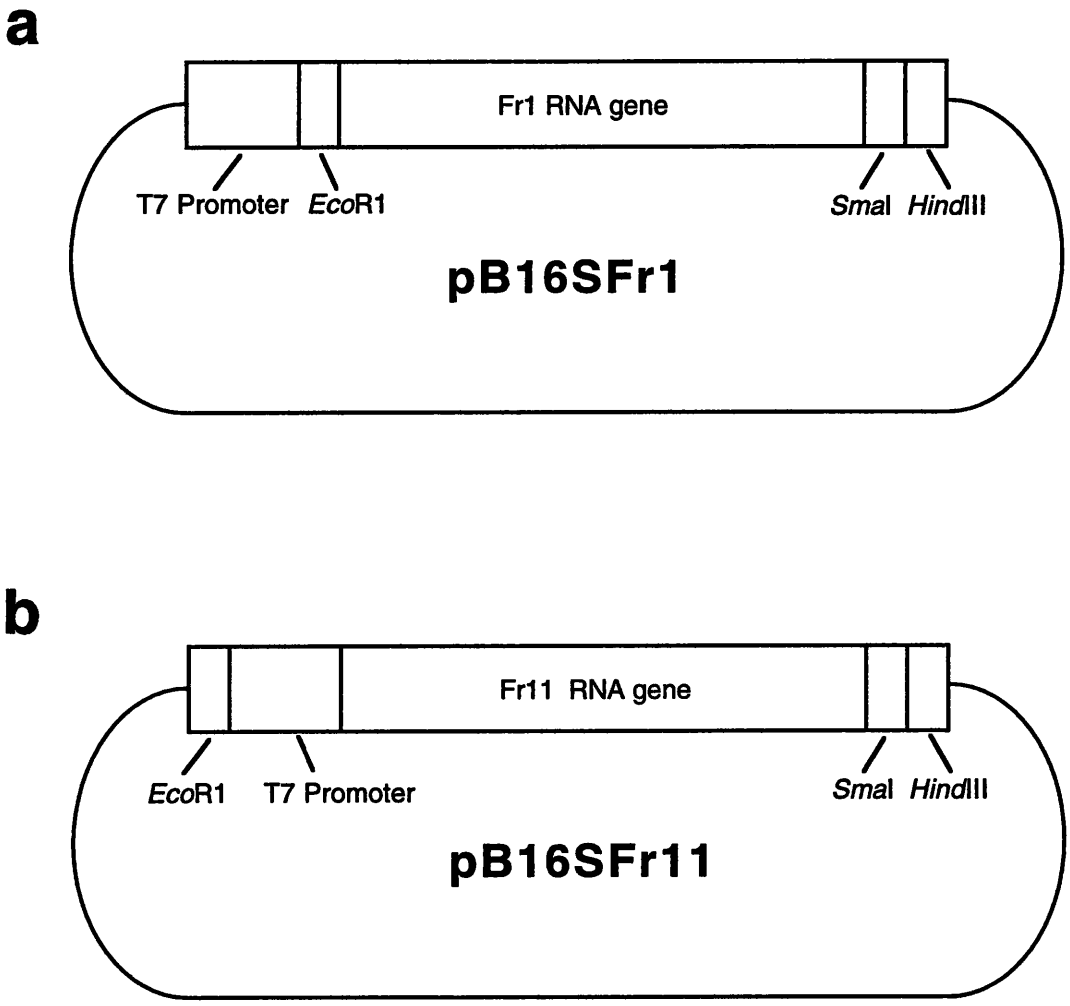


Figure A1.5: Plasmid constructs containing RNA genes. (a) The vector pB16SFr1 and its derivatives, pB16SFr2 through pB16SFr9, were all cloned by inserting the PCR-constructed RNA gene into the vector pGEM-3Zf(-) (Promega) at the *Eco*RI and *Hind*III sites in the multiple cloning site. pGEM-3Zf(-) contains the T7 RNA polymerase promoter immediately upstream of the *Eco*R1 site. (b) The vector pB16SFr11. The constructs pB16SFr10 through pB16SFr17 were constructed by inserting the constructed RNA gene into the vector pUC19 using the *Eco*RI and *Hind*III sites in the multiple cloning site. Since pUC19 has no T7 RNA polymerase promoter, one was placed immediately downstream of the *Eco*R1 site. In all cases, a *Sma*I site was placed upstream of the *Hind*III site for cleavage of the plasmid for run-off transcription.

and an antisense primer that contained sequence complementary to the 5' end of the downstream segment and a sequence encoding a stable RNA tetraloop to bridge the deletion, $^{5'}$GCTCGCGTCAGGTGCAGGCCAGAGAGCCAAAGCTCTCCTCTCCT-GCCCTCAAGTCCC. The downstream segment of the Fr1 gene was amplified using a sense primer containing a sequence complementary to the 3' end of the upstream segment and the same stable tetraloop to bridge the deletion, $^{5'}$GGGACTTGAGGG-CAGGAGAGGAGAGCTTTGGCTCTCTGGCCTGCACCTGACGCGAGC, and an antisense primer complementary to the 3' end of the Fr1 gene containing a *Hind*III site for cloning and a *Sma*I site for linearization of the plasmid for runoff transcription, $^{5'}$CCGCCAAGCTTCCCGGGCGACTTCAAGCTCGCGTCAGGTGCAGGCCAGAGA GC. Amplification was carried out using the same procedure as above, using pB16SFr1 as template DNA. A 5 µl aliquot was subjected to electrophoresis on a 8% (w/v) native polyacrylamide gel in TBE buffer, and the products were cut out and eluted into 100 µl of TE buffer. A second PCR amplification reaction was performed using the 5' upstream primer and the 3' downstream primer and 1 µl of each of the eluted upstream and downstream segments amplified in the first PCR reaction. The product of the second PCR reaction was inserted into pGEM3zf(-) as described above, to generate pB16SFr4, which encodes the Fr4 RNA. This approach was used to also generate plasmids encoding the Fr5, Fr6, and Fr7 RNAs.

The plasmids encoding RNAs of less than 100 nucleotides were synthesized as two overlapping oligonucleotides which were subsequently extended using AmpliTaq DNA polymerase. The insert for pB16SFr11 was generated using a sense oligonucleotide containing an *Eco*RI restriction site and the T7 RNA polymerase promoter, $^{5'}$GCGCCCGAATTCTAATACGACTCACTATAGGGTTGCGGTTCCGAAAGGAACTT GAGGGCAGGAGAGGAGAGCTTC, and an antisense primer containing a *Hind*III site for cloning and a *Sma*I site for linearizing the plasmid, $^{5'}$CGGGCCAAG-CTTCCCGGGTTGCGTCAGGTGCAGGCCAGAGAGCCGAAGCTCTCCTCTCCTG-

CCCTCAA. The double stranded insert was generated in a standard PCR reaction, except that the reaction was only cycled 10 times. The resulting fragment was cloned into pUC18 cleaved with *Eco*RI and *Hind*III using the procedures described above. This method was used to also generate the plasmids encoding the RNAs Fr10 through Fr17, and the structure of pUC18 derived plasmids is shown in Figure A1.5b. The plasmids encoding Fr8 and Fr9 were also generated in this fashion, except that the inserts were cloned into pGEM3zf(-).

Preparation of Fr1 RNA mutants

Mutations were introduced into pB16SFr1, a plasmid containing the gene for expression of Fr1 RNA, using recombinant PCR (Innis *et al.*, 1990). Mutagenic primers were synthesized on an Applied Biosystems PCR Mate 391 DNA Synthesizer. The upstream fragment was amplified using the 5' primer for the Fr1 gene and a mutagenic primer $^{5'}$TCCTGCCCTCAGTCCCCCAG, and the downstream fragment was amplified using a 3' primer for the Fr1 gene, $^{5'}$GCCGCCAAGCTTCCCGGGCGA-ATTCAAACTCGCG and the mutagenic primer $^{5'}$CTGGGGGACTGAGGGCAGGA. The PCR reactions were carried out using the previously described conditions and using pB16SFr1 as template DNA. After purification of the products on an 8%, 29:1 acrylamide:bisacrylamide gel under native conditions, a second PCR reaction was performed using an equal mixture of the product DNA of the first PCR reaction and the 5' and 3' Fr1 gene primers to ligate the two fragments together. Cloning, purification, and sequencing of the mutant plasmids were carried out using the techniques described above.

Synthesis and purification of RNA substrates

Plasmids pB16SFr1 and its derivatives were linearized by cleaving at the unique *Sma*I site engineered at the 3' terminus of the RNA-encoding inserts. To prepare sufficient

quantities of this plasmid for large-scale transcription reactions (45-60 ml volumes), five liters of pB16SFr15/JM109 (pB16SFr15 is a high copy plasmid, derived from pUC18) were grown in Luria Broth, and the plasmid was purified using two Giga-Prep (QIAGEN) plasmid purification columns (each column has a capacity of 10 mg of plasmid). It was determined that only Qiagen plasmid purification method is capable of yielding plasmid of sufficient purity on a large scale; other plasmid purification kits often yielded smaller nucleic acid contaminants of 75-100 nucleotides in size, as observed on an agarose gel. Linearization of the plasmid for run-off transcription at the unique *Sma*I site was performed in 500 µl reactions containing 5 mg of plasmid and 200 U of *Sma*I (New England Biolabs), and incubated for 24 hours at 25 °C. The extent of cleavage was monitored by running 1 µl aliquots on a 1% TAE agarose gel along with uncut plasmid as a control. After 24 hours, another 100 U of enzyme was added, and incubation was continued for another 12 hours. This was repeated until all of the plasmid had been cleaved, as determined using gel electrophoresis (typically, this required 300-400 U of *Sma*I). After the linearized plasmid DNA was phenol:chloroform extracted and ethanol precipitated, it was suitable for transcription reactions.

Transcription reactions contained the following components: 40 mM Tris-Cl, pH 8.1, 1 mM spermidine, 5 mM DTT, 0.01% Triton X-100, 50 mM $MgCl_2$, 8 mM each ribonucleotide 5' triphosphate, 200 µg/ml *Sma*I linearized plasmid, 0.5 µg/ml inorganic pyrophosphatase and 3000 U/ml T7 RNA polymerase (Milligan *et al.*, 1987). The reactions were incubated for 4 hours at 37 °C, and quenched by extracting with an equal volume of phenol equilibrated with TE buffer, pH 8.0. The upper aqueous layer was removed and ethanol precipitated by adding 0.1 volume 3 M sodium acetate, pH 5.2, and 2.5 volumes of 100% ethanol. The precipitate was resuspended in Formamide Stop buffer and purified by electrophoresis on a 8% (w/v) denaturing polyacrylamide gel in TBE buffer. The product band was detected by UV shadowing the gel, excised, and electroeluted using an Elutrap (Schleicher & Schuell) into TBE buffer for 3 hours at 200 V

at 4 °C. The RNA was precipitated with ethanol, and the precipitate was washed once with 70% ethanol prior to resuspension in TE buffer, pH 8.0. The concentration of RNA was assayed by hydrolyzing the product to completion at 37 °C for 24 hours in 0.8 M NaOH and determining the UV absorbance of the hydrolyzed RNA at 260 nm at a neutral pH. The extinction coefficient was calculated for each RNA as the sum contributions of each nucleotide in the RNA sequence.

To label RNA transcripts at the 5' terminus, dephosphorylated RNAs were incubated in the presence of [γ-^{32}P] ATP (ICN, 7000 Ci/mmol), T4 polynucleotide kinase and T4 polynucleotide kinase buffer supplied with the enzyme (New England Biolabs), and 40 units RNasin ribonuclease inhibitor (Promega) at 37 °C for 30 minutes. The reaction was terminated by adding an equal volume of Formamide Stop buffer and purified on a 8% denaturing polyacrylamide gel in TBE buffer. Labeled transcripts were visualized by autoradiography, excised from the gel and soaked for 2 hours in 300 µl 0.5 M sodium acetate, pH 5.2, 0.1% (w/v) SDS, and 1 mM EDTA at 4 °C. The eluted RNA was precipitated with 2.5 volumes of 100% ethanol at -20 °C. The RNA was resuspended in TE buffer and quantitated by Cerenkov counting. RNA transcripts were labeled at the 3' end using [5'-^{32}P] cytidine 3',5' bisphosphate and T4 RNA ligase, and 40 units RNasin ribonuclease inhibitor (England & Uhlenbeck, 1978), and purified as above.

To determine the concentration of 5' or 3' labeled RNA, a separate denaturing polyacrylamide gel was run with a small fraction of the labeling reaction. The gel was imaged using a Molecular Dynamics Phosphorimager, and the quantity of unincorporated label, total labeled RNA (there was always a 20-40% fraction of RNA which had been degraded during the labeling procedure), and labeled product RNA was determined. Using the known quantity of RNA added to the labeling reaction and the specific activity of the label, the concentration of RNA in the eluted sample was calculated.

Synthesis of RNA from DNA oligonucleotides templates

For RNAs shorter than forty nucleotides (*i.e.*, Fr18 RNA), it becomes practical to use single-stranded DNA templates for transcription rather than plasmid templates. This approach utilizes a transcription template composed of two single-stranded DNA oligonucleotides annealed to form a single-stranded template region and a double-stranded promoter region for T7 RNA polymerase directed run-off transcription (Figure A1.6). Transcription reactions contained the following components: 40 mM Tris-Cl, pH 8.1, 1 mM spermidine, 5 mM DTT, 0.01% Triton X-100, 50 mM $MgCl_2$, 80 mg/ml polyethylene glycol (8000 molecular weight), 4 mM each ribonucleotide 5' triphosphate, 300 nM each DNA strand, 0.5 μg/ml inorganic pyrophosphatase and 3000 U/ml T7 RNA polymerase (Milligan *et al.*, 1987).

The are several differences between transcription reaction using plasmid templates and single stranded oligonucleotide templates. Oligonucleotide-directed transcription requires the presence of 80 mg/ml PEG-8000 in order to produce high quantities of RNA, whereas this component is not required for efficient transcription off of plasmid templates. Also, transcription from oligonucleotide templates achieves optimal yields at lower NTP concentrations. The distribution of RNA products in these two types of transcription reactions also differs significantly. With oligonucleotide-directed transcription, a significant number of abortive products are observed that correspond to RNAs that failed to be transcribed completely by the polymerase. Transcription from plasmid templates using T7 RNA polymerase tends to be more processive, and fewer abortive products are observed. However, both templates appear to suffer equally from T7-directed addition of a single nucleotide to the 3' end of the RNA transcript, a significant problem for many RNAs.

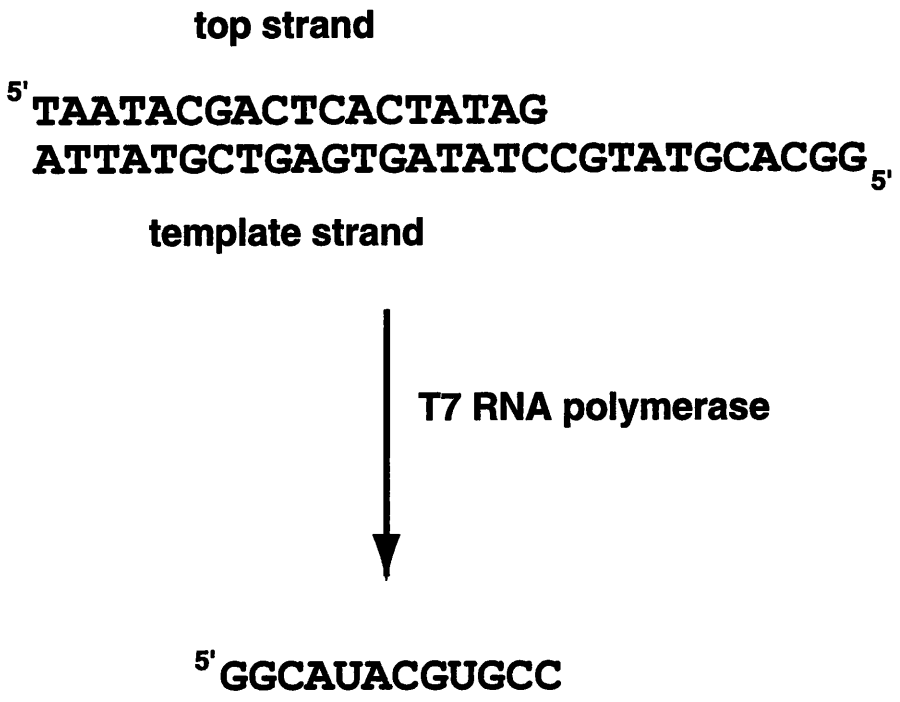


Figure A1.6: Transcription of RNA oligonucleotides from synthetic DNA oligonucleotide templates annealed to form a double-stranded promoter region and a single-stranded template region. Transcription of RNA is intiated from the last base pair of the promoter region.

Appendix 2: Preparation of RNAs labeled with stable nuclei for NMR analysis

Introduction

In order to address the structure of large RNA and RNA-protein complexes using nuclear magnetic resonance (NMR) spectroscopy, it was necessary to develop new tools that would overcome some of the fundamental limitations of this technique. The most labor intensive aspect of determining RNA structure using NMR spectroscopy is obtaining a complete set of proton resonance assignments. This is made particularly difficult for RNA since the ribose sugar H2', H3', H4', and H5'/5" all resonate between four and five ppm, which for RNAs larger than 30 nucleotides leads to tremendous difficulties making these assignments. Cross peaks that are overlapped in two-dimensional experiments can be resolved by coupling these resonances to those of the ^{13}C and ^{15}N heteronuclei to which they are attached in three- and four-dimensional heteronuclear NMR experiments. Another problem associated with homonuclear NMR spectroscopy is that many of the proton assignments are made using nOe experiments, and thus dependent upon the conformation of the macromolecule. Procedures for making unambiguous assignments for base and sugar protons instead take advantage of experiments which exploit the efficient through-bond transfer of magnetization using the large ^{1}H-^{13}C (145-160 Hz) and ^{13}C-^{13}C (38-45 Hz) one-bond couplings.

Recent advances in structure determination using NMR spectroscopy have relied heavily on uniform isotopic labeling to permit application of higher dimensional heteronuclear NMR techniques (Ikura *et al.*, 1990; Clore & Gronenborn, 1991; Clore *et al.*,

1991). Use of 3D- and 4D-NMR methodology in conjunction with ^{15}N- and ^{13}C-labeling has made a significant impact on protein solution structure determination. Proteins are usually prepared by overexpression in bacteria, and isotopic labeling only requires growth in isotopically labeled media. These techniques have only recently been applied to nucleic acids (Nikonowicz & Pardi, 1992b; Nikonowicz & Pardi, 1992a; Santoro & King, 1992). DNA and RNA are routinely prepared either by chemical or enzymatic synthesis, and thus, isotopic labeling of nucleic acids requires the preparation of isotopically labeled monomers in a suitable form.

This chapter describes a procedure for preparation of uniformly isotopically labeled ribonucleoside triphosphates and their utilization in large scale *in vitro* transcription reactions to prepare isotopically labeled RNAs for NMR study. A similar method has been used to prepare uniformly ^{13}C-labeled ATP for mass spectral studies (Polson *et al.*, 1991). The procedure, outlined in Figure A2.1, is general and can be adapted to the preparation of ^{13}C- and/or ^{15}N-labeled RNA. Bacterial cells are grown in a minimal salts medium containing isotopically labeled carbon and/or nitrogen substrates. *E. coli* are grown to produce ^{15}N-labeled nucleotides, whereas ^{2}H labeled, ^{13}C-labeled, or ^{13}C/^{15}N nucleotides are best produced by growing the obligate methylotrophic bacterium *Methylophilus methylotrophus* on ^{13}C-methanol, which is an economical source of ^{13}C. The cells are harvested and lysed by detergent, the proteins are removed by phenol-chloroform extraction, and the total nucleic acids are precipitated with isopropanol. The total cellular nucleic acids are hydrolyzed to nucleoside monophosphates using nuclease P1, and the deoxyribonucleotides separated from the ribonucleotides using a boronate affinity column. The ribonucleoside monophosphates are then enzymatically converted to nucleoside triphosphates, which can be used in transcription reactions to prepare any desired RNA of defined sequence.

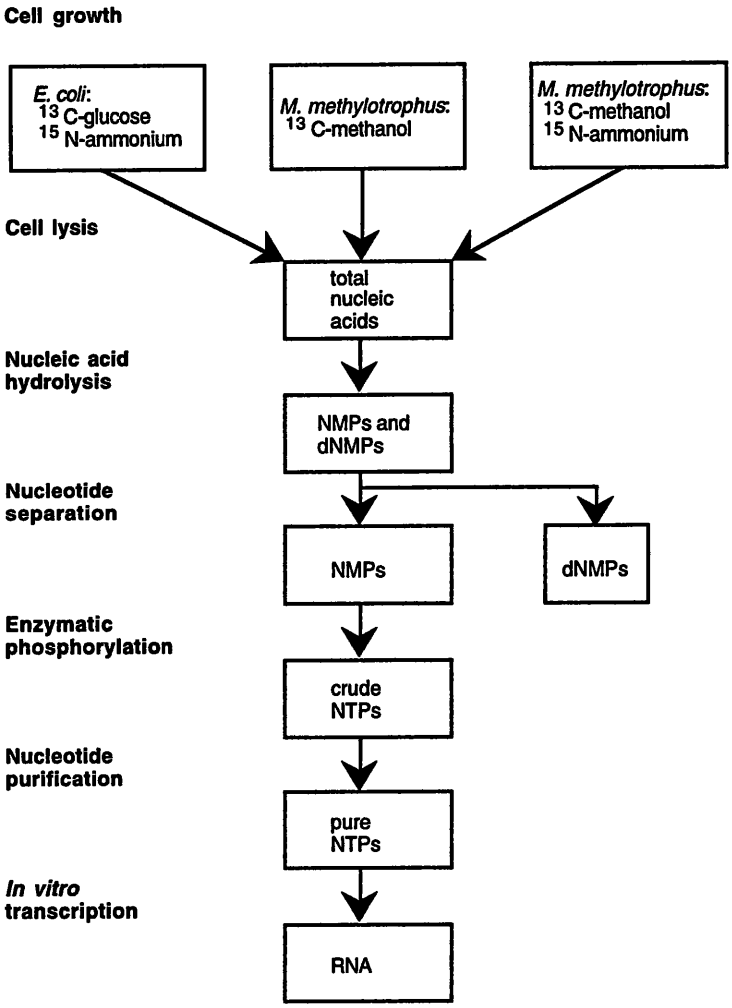


Figure A2.1: Flowchart outlining the procedure for generating isotopically labeled nucleoside triphoshphates.

Methodology

Growth of E. coli

For the synthesis of ^{15}N-labeled nucleic acids, *E. coli* (ATCC (American Type Culture Collection) 15244) is the organism of choice because it is most easily cultured. However, for routine synthesis of nucleic acids incorporating ^{13}C, the procedure described here that utilizes ^{13}C-glucose as a carbon source is prohibitively expensive, and thus was only performed once to compare the yields with that of *Methylophilus methylotrophus.*

For ^{15}N-labeled nucleic acids, *E. coli* (ATCC 15224) was grown on the glycerol-minimal medium described in Table A2.1. A solution containing glycerol and the phosphate salts was made separately, the pH adjusted to 7.0, and autoclaved. Sodium citrate, ammonium sulfate, and magnesium sulfate were made as individual 100x solutions and autoclaved to sterilize. A 1000x trace metals solution was made as a mixture of all of the individual metals and filter sterilized. The other nutrients were added after the autoclaved media had cooled.

A five ml culture of *E. coli* was grown on glycerol-minimal salts medium containing ^{14}N-ammonium sulfate overnight at 37 °C with agitation. One ml of this culture was used to inoculate a one liter culture of glycerol-minimal salts/phosphate medium containing ^{15}N-ammonium sulfate in a 4 L Erlenmeyer flask and incubated at 37 °C with vigorous agitation for 15 hours. The cells were harvested by centrifugation at 6400 g (6000 rpm in a JA-10 rotor (Beckman)) for 20 minutes. Each liter of cell culture yielded approximately 4 g of wet packed cells.

To obtain ^{13}C and ^{15}N-labeled nucleic acids, *E. coli* was grown on the glucose-minimal medium described in Table A2.1. A solution containing only KH_2PO_4 was made

separately, the pH adjusted to 7.0 with KOH, and autoclaved. The other components were then added as filter sterilized solutions, along with five ml of a filter sterilized 20% (w/v) solution 98% ^{13}C-glucose (Cambridge Isotope Laboratories). Subsequent growth is the same a described above.

Growth of Methylophilus methylotrophus

The ATCC lyophilized stock of *Methylophilus methylotrophus* was revived in 5 ml of ATCC Medium 1545 with shaking at 30 °C for 2-3 days until the cultures become very cloudy. The revived culture was plated out on ATCC Medium 1545 plates and incubated for 2 days at 37° C. ATCC Medium 1545 (Table A2.1) was prepared as a one liter solution containing only the phosphate salts with the pH adjusted 6.8, and then autoclaved. Individual 100x stock solutions of ammonium chloride, calcium chloride, and magnesium chloride were made and autoclaved separately. A 1000x trace metals solution was made as a mixture of all of the individual metals and filter sterilized. After the phosphate solution had cooled, these solutions, as well as the methanol, were added. For plate medium, 15 grams of agar was added to the phosphate solution prior to autoclaving.

Methanol-minimal salts medium (Table A2.1) (Beardsmore *et al.*, 1982) was prepared by autoclaving one liter of a solution containing only the phosphate salts, with the pH adjusted to 6.8. Ammonium sulfate and magnesium sulfate were made as individual 100x solutions, and autoclaved separately. A 1000x trace metals solution was made as a mixture of all of the individual metals and filter sterilized. After the solution had cooled, these solutions, along with methanol, were added. A 5 ml overnight culture in methanol-minimal salts medium was inoculated with an individual colony, and grown for 12 to 16

Table A2.1: Media utilized for growth of bacteria for incorporation of stable isotopes.

Glycerol-Minimal Salts Medium (per liter):

glycerol	21.5 g
KH_2PO_4	1.6 g
NaH_2PO_4	5.3 g
Na·citrate	0.5 g
$(NH_4)_2SO_4$	0.7 g
$MgSO_4 \cdot 7H_2O$	0.3 g
Trace Metals	
$CaCl_2 \cdot 2H_2O$	750 mg
Na_2EDTA	30 mg
$FeCl_3 \cdot 6H_2O$	25 mg
$CuSO_4 \cdot 5H_2O$	240 µg
$MnSO_4 \cdot 5H_2O$	180 µg
$ZnSO_4 \cdot 7H_2O$,	27 µg
$CoCl_2$	270 µg

Glucose-Minimal Salts Medium (per liter):

glucose	5 ml (20% glucose soln, w/v)
KH_2PO_4	13.6 g
$(NH4)_2SO_4$	0.3 g
$MgSO_4 \cdot 7H_2O$	0.25 g
$CaCl_2 \cdot 2H_2O$	15 mg
Trace Metals	
Na_2EDTA	30 mg
$FeCl_3 \cdot 6H_2O$	25 mg
$CuSO_4 \cdot 5H_2O$	240 µg
$MnSO_4 \cdot 5H_2O$	180 µg
$ZnSO_4 \cdot 7H_2O$,	27 µg
$CoCl_2$	270 µg

Table A2.1: Continued.

ATCC Medium #1545 (per liter):

CH_3OH	5 ml
Na_2HPO_4	0.33g
KH_2PO_4	0.26 g
NH_4Cl	0.5 g
$CaCl_2$	0.2 g
$MgSO_4 \cdot 7H_2O$	1.0 g
Trace Metals	
Fe EDTA	5.0 mg
$FeSO_4 \cdot 7H_2O$	500 µg
$NaMoO_4 \cdot 2H_2O$	2.0 mg
$ZnSO_4 \cdot 7H_2O$	400 µg
$MnCl2 \cdot 4H_2O$	20 µg
H_3BO_3	15 µg
$CoCl_2 \cdot 6H_2O$	50 µg
$NiCl_2 \cdot 6H_2O$	10 µg
EDTA	250 µg

Methanol-Minimal Salts Medium (per liter):

CH_3OH	1 ml
K_2HPO_4	0.95 g
NaH_2PO_4	0.78 g
$(NH_4)_2SO_4$	0.36 g
$MgSO_4 \cdot 7H_2O$	0.2 g
Trace Metals	
$FeSO_4 \cdot 7H_2O$	50 mg
$CuSO_4 \cdot 5H_2O$	100 µg
$MnSO_4 \cdot 5H_2O$	50 µg
$ZnSO_4 \cdot 7H_2O$	50 µg
$CaCl_2 \cdot 2H_2O$	1.3 mg
$CoCl_2$	10 µg
H_3BO_3	7 µg
$NaMoO_4$	10 µg

hours with shaking at 37 °C. Growth in isotopically labeled medium was performed in three stages: an unlabeled 50 ml culture, a labeled 500 ml culture, and a labeled 10 L culture. At all stages during this procedure, it is imperative to keep the cultures actively growing and not allow them to reach stationary phase. Growth was monitored by reading the optical density at 660 nm. A 50 ml culture containing ^{12}C-methanol-minimal salts medium was inoculated with 1 ml of the overnight liquid culture stock, and grown with vigorous shaking at 37 °C for approximately 12 hours. A 500 ml culture of ^{13}C-methanol-minimal salts medium was inoculated with one ml of the 50 ml ^{12}C-culture to minimize isotopic dilution, and incubated at 37 °C with agitation for approximately 18 hours. The entire 500 ml ^{13}C-culture was used to inoculate 10 L of ^{13}C-methanol-minimal salts medium in a microfermentor (New Brunswick). The microfermentor was set to 37 °C, 5 L/min air flow rate, and 600 rpm rotor speed for agitation. Growth was monitored by taking 660 nm absorbance readings every hour (Figure A2.2). When the cell density of the culture reached an absorbance of 0.4 - 0.45, the growth of the cells tapered off, indicating the transition to stationary phase and the temperature of the fermentor was turned down to 4 °C. The culture was harvested by centrifugation of the culture in six 500 ml centrifuge bottles at 6400 g (6000 rpm in a JA-10 rotor) for 20 minutes. This yields 15-16 grams of pink, tightly-packed wet cell pellets distributed between six 500 ml centrifuge bottles. During centrifugation, the remaining culture continued to be agitated in the fermentor at 4 °C. After each centrifugation, the supernatant was poured off, and more culture was added to each bottle. Four centrifuge runs were required to harvest 10 L of culture.

For random fractional deuteration of ribonucleotides, it was important to obtain high density cultures, without compromising the yield of nucleic acids, because large volumes of D_2O are expensive and time consuming to recycle. Thus, in order to obtain preparative quantities of labeled ribonucleotides for NMR studies, the protocol for culturing *M. methylotrophus* was modified to yield high density cultures, by controlling the pH of

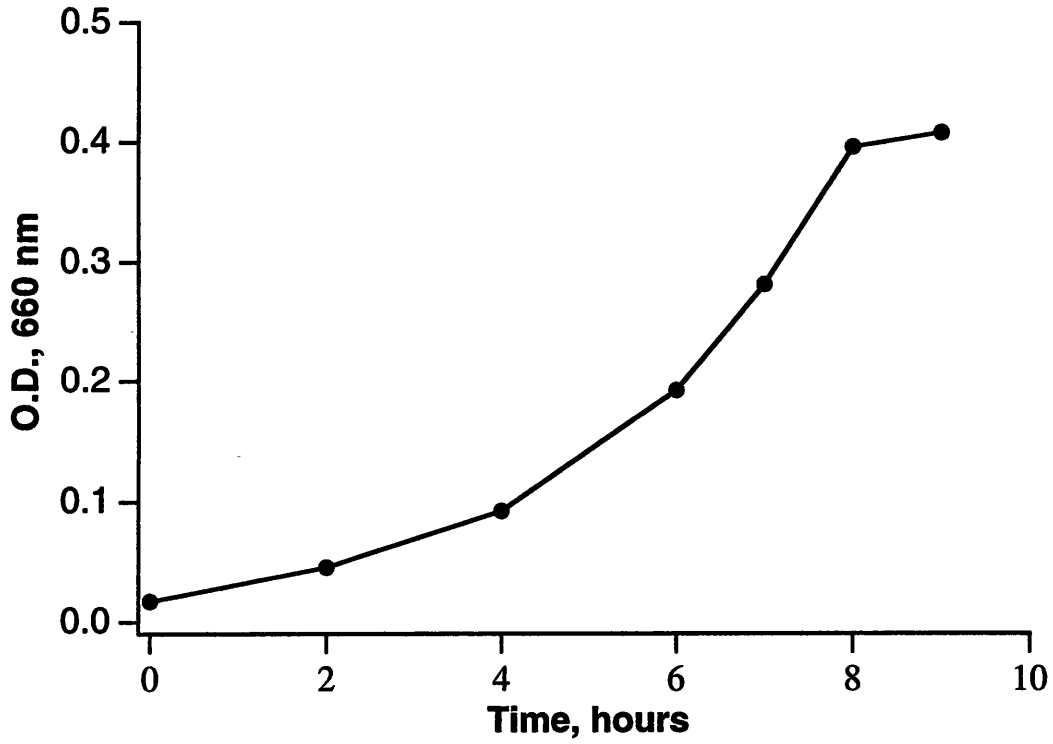


Figure A2.2: Growth curve of *Methylophilus methylotrophus* in a 10 L fermentor. Culture was grown in methanol-minimal salts medium with an air flow rate of 5 L/min, agitation speed of 600 rpm, and at 37°C. Absorbance readings were taken at 600 nm. Using the absorbance readings from 2 to 8 hours, the doubling time of the culture is 1.9 hours.

the medium during growth. The methanol-minimal salts medium used for this application is the same as described previously, except that the medium was prepared using a solution of 52:48 ^2H:^1H water. The ^2H:^1H ratio of the medium was determined by integrating the NMR spectrum of a mixture of 200 µl of dioxane and 600 µl of medium. Cultures of *M. methylotrophus* grown in H_2O exhibit a doubling time of 1.9 hours, which is increased to 4.0 hours for growth in media containing 60% D_2O. However, during successive serial passages of these cells into fresh medium containing 60% D_2O, the doubling time dropped to 2.5 hours Acclimation of the cells to D_2O prior to large scale growths was critical for ensuring high nucleic acid yields, with a favorable RNA:DNA ratio (Herbert, 1961).

The initial phase of growing *M. methylotrophus* in the 10 liter culture was performed exactly as described above. When the optical density (O.D.) of the culture at 660 nm approached 0.4, the pH, measured by a standard pH probe inserted into the culture, had typically dropped to 6.1-6.2. The pH of the culture was then adjusted to 6.6 by addition of 0.5 M NaOH/KOH (1:1 molar ratio) in 52:48 D_2O/H_2O to the culture and maintained at that pH for the duration of the growth. When the culture reaches an O.D. of 0.4 at 660 nm, nearly all of the methanol and ammonium sulfate in the medium has been exhausted by the bacteria. The growth is continued by addition of 10 ml of methanol, 3.6 g ammonium sulfate, and 11.8 ml 100% D_2O to maintain the 52:48 ^2H:^1H ratio in the media. After approximately two hours, the cell density in culture had doubled, and a second addition of methanol, ammonium sulfate, and D_2O was added to the culture, and the pH was again adjusted to 6.6. A third addition was repeated after approximately 1.5 hours when the O.D. of the culture reached 1.3, and the fourth addition after another hour, when the O.D. of the culture reached 1.8. After the final aliquot of nutrients was added, the culture was allowed to continue to grow until the O.D. reached 2.2-2.5 (a typical growth curve is shown in Figure A2.3). The growth was terminated by reducing the temperature of the culture to 4 °C, and harvesting the cells by centrifugation at 8000 rpm in

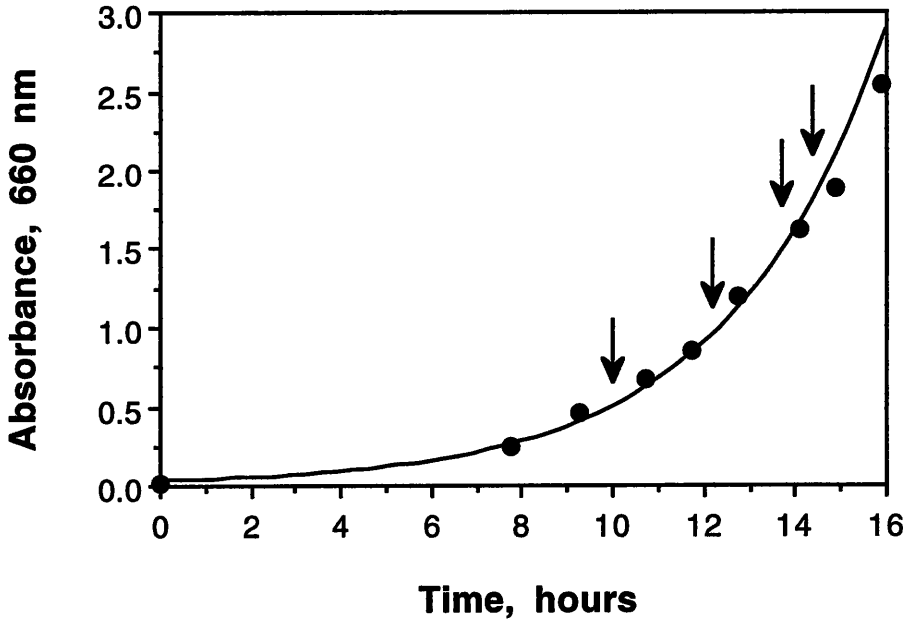


Figure A2.3: Growth curve of a 10 L culture of *M. methylotrophus* grown in media containing a 52:48 ^{2}H:^{1}H ratio. The initial doubling time of the culture was 2.4 hours, and gradually increased towards the end of the growth. The arrows indicate the times at which additional nutrients were added to the culture; the pH of the culture was adjusted to 6.6 whenever it dropped below 6.4.

a JA-10 rotor for 15 minutes, yielding 70-80 grams of wet packed cells. After each growth, the D_2O/H_2O mixture from spent media was recycled by rotary evaporation (Moore, 1979).

Cellular Lysis

Phenol containing 0.1% 8-hydroxyquinoline was prepared from solid reagent grade phenol (Mallikrodt) by equilibrating against STE buffer (Sambrook *et al.*, 1989). The solid phenol was liquefied by incubating in a 65 °C water bath. The liquid phenol was extracted twice against 0.1 M Tris, pH 8.0 and twice against STE buffer in a one liter separatory funnel. 0.1% w/v 8-hydroxyquinoline was added to the buffered phenol and the phenol solution was stored in an amber bottle at 4 °C.

Approximately 4 g of wet packed cells were resuspended in an equal volume of STE buffer (0.1 M NaCl, 10 mM Tris·Cl, pH 8.0, 1 mM EDTA, pH 8.0). Resuspended cells were slowly added to a mixture of 90 ml of STE buffer and 5 ml 10% SDS at 37 °C with rapid stirring. Phenol containing 0.1% 8-hydroxyquinoline (w/v) was equilibrated against STE buffer, a 24:1 chloroform:isoamyl alcohol solution was prepared, and a 1:1 solution of the phenol and chloroform solutions was preheated to 65 °C. After the cellular lysate had been thoroughly homogenized for 10-15 minutes, the solution was added to 200 ml of preheated phenol:chloroform and incubated for 30 minutes at 65 °C with occasional vigorous mixing to keep the organic and aqueous phases emulsified. The aqueous and organic phases in the emulsion were separated by centrifuging for 10 minutes at 6400 g (6000 rpm in a JA-10 rotor) in 500 ml polypropylene (or other chemically compatible) centrifuge bottles. The upper aqueous layer was removed with an aspirator, taking care not to disrupt the white protein inclusion layer. The organic and inclusion layers were extracted twice with 100 ml of STE buffer in a Waring blender set on high for 5 minutes, and the aqueous layer was removed after centrifugation.

The aqueous layers from each extraction were pooled (~300 ml total volume) and extracted with an equal volume of 24:1 chloroform:isoamyl alcohol. The phases were allowed to completely separate without centrifugation before the aqueous layer was removed. Nucleic acids were precipitated by adding 30 ml 3 M sodium acetate, pH 5.2, and 330 ml isopropyl alcohol to the aqueous phase and incubating overnight at -20 °C. Precipitation at low temperature for more than four hours is critical for ensuring complete precipitation of all cellular nucleic acids. The solution was centrifuged at 6400 g (6000 rpm) for 30 minutes in 500 ml centrifuge bottles in a JA-10 rotor, the liquid decanted, then the pellet was air-dried and slowly resuspended in 25 ml of STE buffer. Four grams of wet cells should yield approximately 3500 A_{260} units of crude nucleic acids for both *Methylophilus methylotrophus* and *E. coli.*

Nucleic acid hydrolysis

The resuspended nucleic acids from a single lysis procedure (3250-3500 A_{260} units in 25 ml STE) were denatured in a boiling waterbath for 5-6 minutes. The solution was cooled in an ice bath before adjusting to 15 mM sodium acetate, and 1.0 mM $ZnSO_4$, pH 5.2 by adding 125 µl 3 M sodium acetate, pH 5.2 and 625 µl 40 mM $ZnSO_4$. Nucleic acids were hydrolyzed by adding 6 units nuclease P1 and incubating for 1 hour at 37 °C (Haynie & Whitesides, 1990). To ensure complete hydrolysis of the nucleic acids, another 6 units of nuclease was added after one hour, and incubation was continued for an additional hour.

Reverse phase HPLC was used to monitor the hydrolysis of DNA and RNA during the P1 nuclease digestion. For efficient analytical separation of ribonucleotides and deoxyribonucleotides, the nucleotides were dephosphorylated with calf intestinal alkaline phosphatase prior to injection. A sample of 20-30 µg of nucleoside monophosphates was dephosphorylated by incubating with 2 units of calf intestinal alkaline phosphatase (Boehringer Mannheim) for 30 minutes at 37 °C in phosphatase buffer (50 mM Tris·HCl,

pH 8.5, 0.1 mM EDTA, pH 8.5), followed by another 2 units of phosphatase and continued incubation for 30 minutes. HPLC analysis of the reaction mixture was performed on a 250x4.6 mm C18 column (Vydac) using an isocratic elution with methanol-83.3 mM triethylammonium phosphate, pH 6.0 (4:96 v/v) as the mobile phase at a flow rate of 1 ml/min (Lim & Peters, 1989). Sample volumes of 10-20 µl were injected and nucleosides were detected at 268 nm with a UV detector, as shown in Figure A2.4.

Nucleotide separation

To prepare the affinity chromatography column, 5 g Affi-gel 601 was hydrated in TE buffer (10 mM Tris·HCl, pH 8.0, 1 mM EDTA, pH 8.0) and packed in a 20x2.5 cm glass column. A wide column was important for good flow rates because the resin volume changes with pH and ionic strength. 1 M TEABC, pH 9.5, was prepared by bubbling CO_2 through 141 ml triethylamine in 700 ml H_2O at 5 °C until the pH drops to 9.5, and bringing to 1 L volume with H_2O. To adjust the pH of 1 M triethylamine to 9.5 required 2 to 4 hours when CO_2 was bubbled through a glass Pasteur pipette. The column was equilibrated with 1 M TEABC at 5 °C. It is crucial that all steps be performed at 5 °C in order for the ribonucleotides to bind to the column.

The nucleotide solution ($\leq$ 6000 A_{260} units) from the P1 digestion was lyophilized and resuspended in 10 ml of 1 M TEABC. The sample was applied to the column, and then washed with 1 M TEABC, while collecting 5 ml fractions, until the A_{260} of the eluant dropped below 0.1 (Figure A2.5) (Schott *et al.*, 1973). The deoxyribonucleotides, salts, and other impurities washed through the boronate column, while the ribonucleotides remained covalently bound to the boronate ligand. To elute the bound material, the column was washed with H_2O that had been acidified to pH 4-5 by bubbling with CO_2. Elution was continued until the A_{260} of the eluant dropped below 0.1, which usually occurred 100 ml after the start of elution. Fractions of low pH eluant >0.1 A_{260} were pooled,

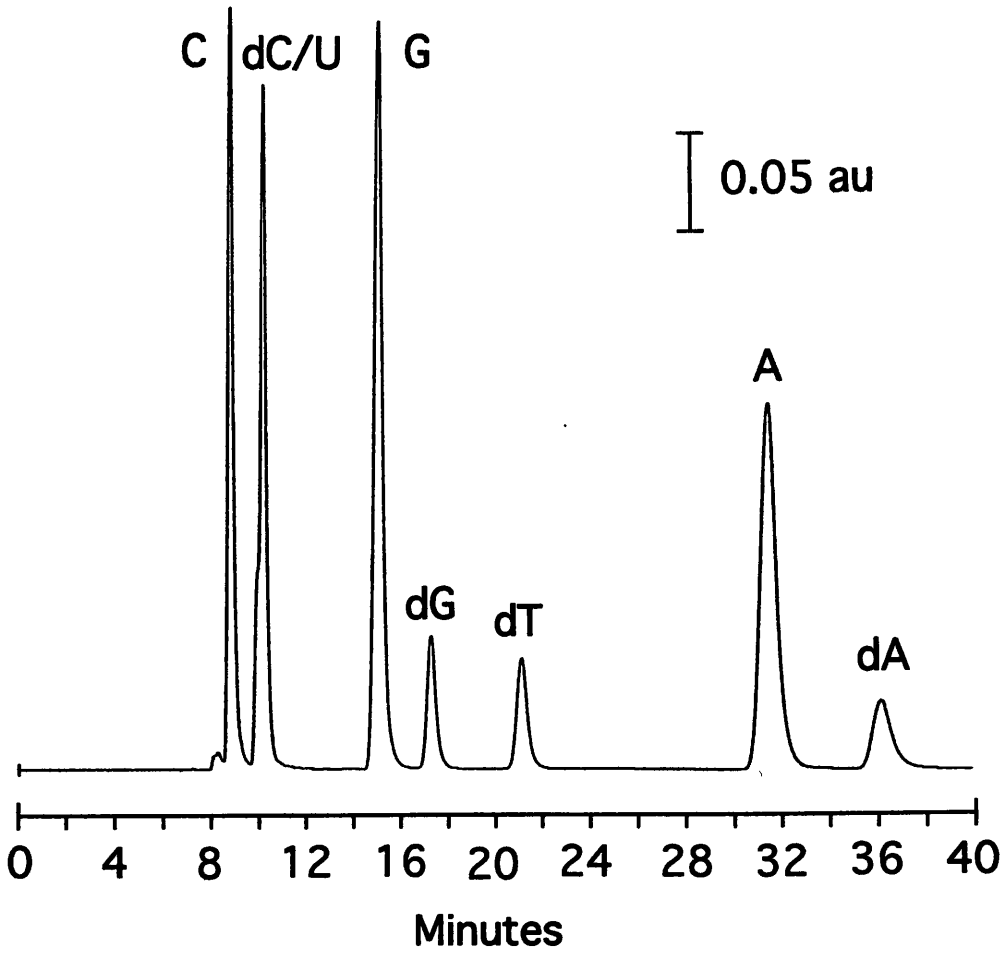


Figure A2.4: Reverse phase HPLC chromatogram of nucleoside mixtures from a culture of *Methylophilus methylotrophus* grown in a 10 L fermentor. A total of 250 µg of nucleoside monophosphates were dephosphorylated with calf intestinal alkaline phosphatase prior to injection onto a Vydac C18 reverse phase column. Elution was isocratic, with methanol-83.3 mM triethylammonium phosphate, pH 6.0 (4:96 v/v), a flow rate of 1.0 ml/min, and detection at 268 nm at 0.5 absorbance units full scale (aufs).

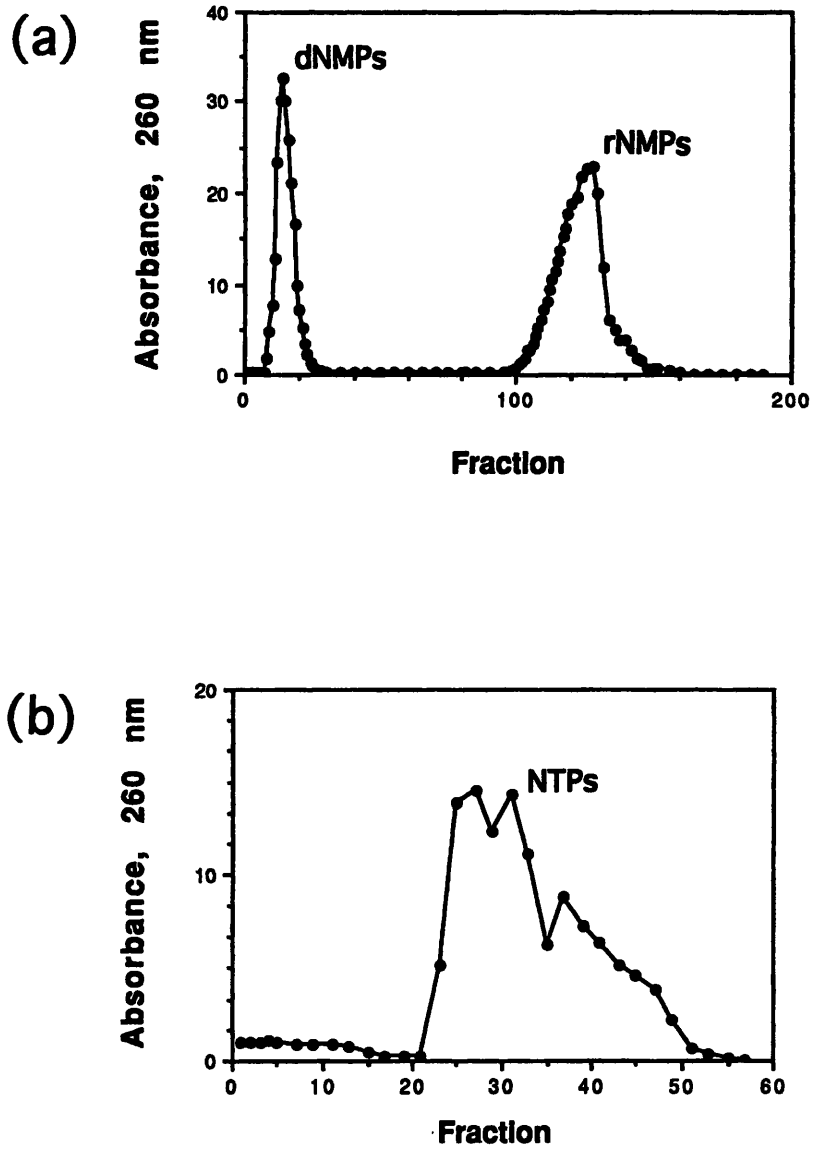


Figure A2.5: Purification of ribonucleotides on an Affi-gel 601 boronate affinity column. The column contained 5 g of hydrated Affi-gel 601 in a 20x2.5 cm glass column, and loading, washing, and elution stepwise was done by gravity. The nucleotides were bound and washed with 1 M TEABC, pH 9.5, and eluted with H_2O that had been acidified to pH 4-5 with CO_2. The absorbance of individual fractions was monitored at 260 nm. (a) Separation of deoxyribonucleotides and ribonucleotides. The total nucleotide pool after P1 nuclease hydrolysis was applied to the boronate affinity column while collecting 2 ml fractions. Elution with acidified water began with fraction 58. (b) Desalting of ribonucleoside triphosphates. NTPs were applied to the boronate affinity column, collecting approximately 7 ml fractions. The NTPs were eluted with acidified water and applied beginning with fraction 13.

lyophilized, and resuspended in 5 ml of H_2O. Complete separation of dNMPs from NMPs using this method was verified by C18 reverse-phase HPLC, as shown in Figure A2.6.

Separation of individual ribonucleotides

The pooled ribonucleotides in 5 ml H_2O were injected onto a 100 x 7.8 mm HP-PEI HPLC column (Interaction Chemicals) in 0.5 ml aliquots containing no more than 70 mg of ribonucleoside monophosphates (~2100 A_{260} units). Separation was carried out using a two step profile: isocratic elution at 100% solvent A (0.05 M NH_4COOH, pH 3.0) for 20 minutes, and then a linear gradient of 0 to 100% solvent B (0.5 M NH_4COOH, pH 2.5) over 20 minutes at a flow rate of 1 ml/min. Nucleotides were detected at 300 nm with a UV detector and 2 minute fractions were collected (typical chromatograms from this separation are shown in Figure A2.7). Fractions corresponding to each nucleotide were pooled and lyophilized to dryness. To completely remove trace amounts of ammonium formate, each pool of nucleotides was resuspended in 20 ml of H_2O and lyophilized to dryness.

Enzymatic phosphorylation

The Na^+ form of AG 50W-X8 (H^+ form, 200-400 mesh) (Bio-Rad) was prepared by treating 25 g of resin with three 50 ml exchanges of 1 M NaCl for 15 minutes for each exchange, and three times with H_2O. The Na^+ form of 3-phosphoglycerate was prepared by vigorously stirring 2 g of Ba^{2+} 3-phosphoglycerate with 15 ml of a 50% slurry of Na^+-AG 50W-X8 resin for 30 minutes. The resin was removed by filtration and washed three times with 5 ml ddH_2O. The pH of the combined filtrates was adjusted to 7.5 using 1 M NaOH. The final concentration of Na^+-exchanged phosphoglycerate was approximately 0.2 M.

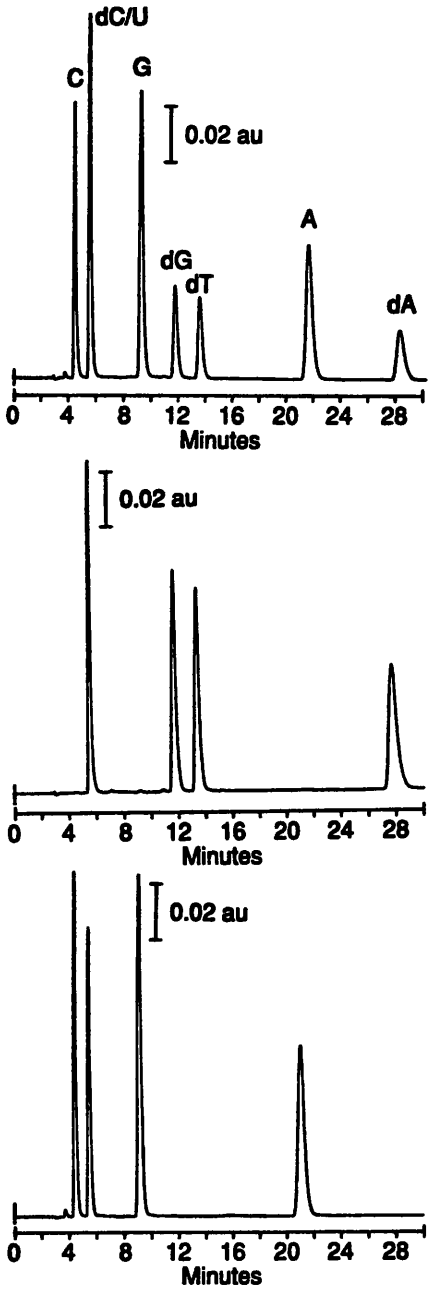


Figure A2.6: Reverse phase HPLC chromatograms of nucleoside mixtures. A total of 50 μg of nucleoside monophosphates were dephosphorylated with calf intestinal alkaline phosphatase prior to injection onto an Econosphere C8 reverse phase column. Elution was isocratic, with methanol-83.3 mM triethylammonium phosphate, pH 6.0 (4:96 v/v), a flow rate of 1.0 ml/min, and detection at 268 nm at 0.2 absorbance units full scale (aufs). (a) Total nucleotide pool from *M. methylotrophus* cellular lysate, prior to boronate affinity chromatography. (b) Deoxyribonucleotides, which do not bind to the boronate column. (c) Bound ribonucleotides, which are eluted from the boronate column.

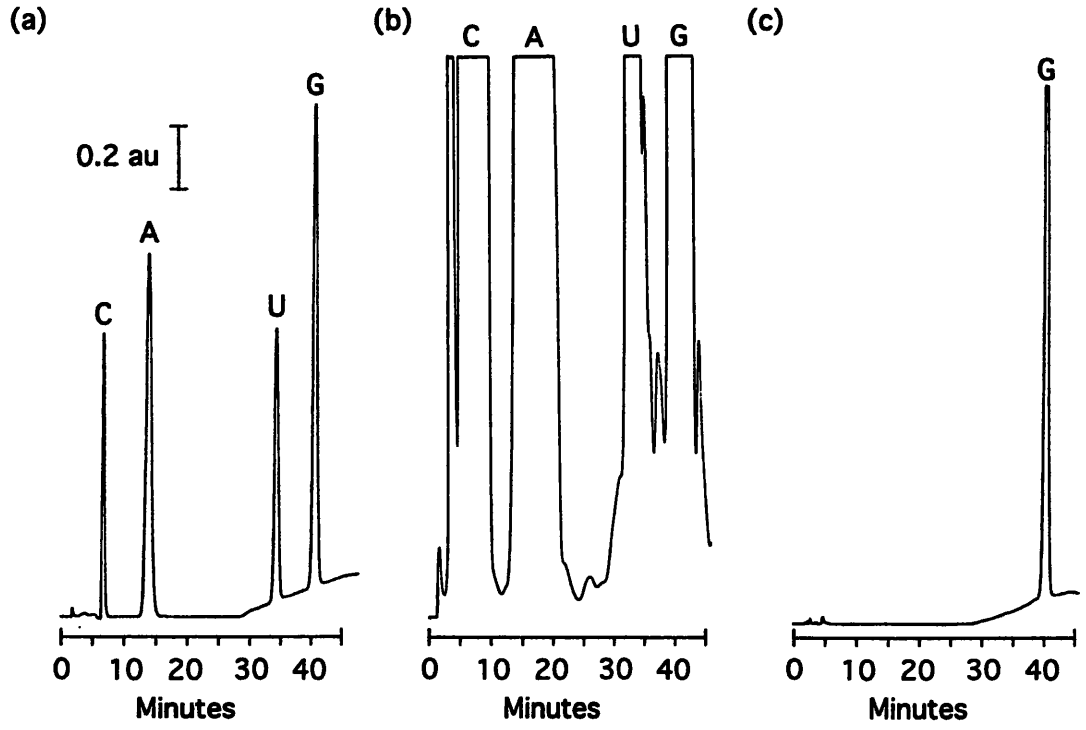


Figure A2.7: Ion exchange HPLC separation of the individual ribonucleoside monophosphates. Ribonucleoside monophosphates were injected onto an HP-PEI ion exchange HPLC column and eluted using the scheme described in the text with a flow rate of 1 ml/min. (a) Injection of approximately 100 µg of total ribonucleoside monophosphates in a volume of 0.5 ml with detection at 268 nm, 2 aufs. (b) Injection of approximately 65 mg of total ribonucleoside monophosphates in a volume of 0.5 ml with detection at 300 nm, 2 aufs. (c) Fractions from injection (b) corresponding to the guanosine monophosphate peak were pooled and 100 µg reinjected in a volume of 0.1 ml, with detection at 268 nm, 2 aufs.

This procedure was derived from a procedure developed by Whitesides (Haynie & Whitesides, 1990). Enzymatic phosphorylation of approximately 83 mg of nucleoside monophosphates to nucleoside triphosphates was performed in a 50 ml three-necked round bottom flask flushed with argon at ambient temperature. The NMPs were added to give a final concentration of 10 mM in a solution containing 15 mM KCl, 75 mM $MgCl_2$, 15 mM dithioerythritol, and 10 mM Na^+ 3-phosphoglycerate (Simon *et al.*, 1990). The pH of the solution was adjusted to 7.5 with 1 M HCl, and was maintained during the course of the reaction with a pH controller (Markson) delivering 0.1 M HCl via a peristaltic pump (Gilson).

The phosphorylation reaction is initiated by first generating ATP. Synthesis of ATP was initiated by adding 1 μM ATP, 10 units phosphoglycerate mutase (from rabbit muscle, Boehringer Mannheim), 200 units myokinase (from chicken muscle, Sigma), 100 units enolase (from baker's yeast, Sigma), and 200 units pyruvate kinase (from rabbit muscle, Sigma). After 3 hours, when ATP represents >90% of the adenosine nucleotide pool, 0.5 units of guanylate kinase (from porcine brain, Sigma), and 1.0 unit nucleoside monophosphate kinase (from beef liver, Boehringer Mannheim) were added, and the concentration of 3-phosphoglycerate was increased to 20 mM. During the synthesis of CTP, GTP, and UTP, the concentration of phosphoglycerate was increased by 5 mM every 3-4 hours. Complete phosphorylation of all nucleotides occurs within 12-14 hours, and the reaction mixture was stored at -20 °C.

This procedure has been used to successfully phosphorylate up to 500 mg of nucleotides by linearly scaling up the volume of the reaction, while the concentration of the individual components was kept constant. The phosphorylation of GMP to GTP was performed using the same protocol as above, with a few modifications. The concentration of catalytic ATP was increased to 0.3 mM; a significant concentration of ATP is required to drive the reaction forward at an appreciable rate. Also, nucleoside monophosphate kinase was omitted from this reaction, while increasing the guanylate kinase added to 3 units.

Conversion of NMPs to NTPs was monitored by HPLC, and several chromatograms during the course of a typical phosphorylation are shown in Figure A2.8. Populations of NMPs, NDPs, and NTPs in were analyzed using a 25x4.6 mm Vydac nucleotide analysis column. Chromatography was carried out using 0.045 M NH_4COOH, pH 4.6 with H_3PO_4 as solvent A and 0.5 M NaH_2PO_4, pH 2.7 with HCOOH as solvent B, with a linear gradient from 0 to 100% solvent B in 10 minutes and a flow rate of 2.0 ml/min. Sample volumes of 20 µl were injected, and nucleotides were detected at 268 nm with a UV detector.

Nucleotide purification

The phosphorylation reaction mixture was lyophilized, resuspended in 10 ml ice-cold 1 M TEABC, pH 9.5, and purified using the boronate affinity column procedure. This desalts the nucleotides, which is critical for their function in transcription reactions *in vitro* . The lyophilized NTPs were resuspended in 2 ml H_2O. Any remaining high molecular weight contaminants were removed by passing the nucleotide solution through a Centricon 10,000 molecular weight cutoff microconcentrator (Amicon) at 2200 g (4000 rpm) in a JA-17 rotor , and collecting the filtrate. The concentration of the NTP solution was estimated by UV absorbance at 260 nm.

In Vitro Transcription of TAR RNA

RNA was synthesized by transcription from an oligonucleotide template with a single-stranded template region and a double-stranded promoter region with T7 RNA polymerase (Milligan *et al.*, 1987). Oligonucleotides were synthesized by standard phosphoramidite chemistry using an automatic DNA synthesizer. A 44 ml transcription reaction was incubated for 4 hours at 37 °C in 40 mM Tris·HCl, pH 8.1, 1 mM spermidine, 5 mM DTT, 9.6 mM $MgCl_2$, 0.01% Triton X-100, 80 mg/ml polyethylene

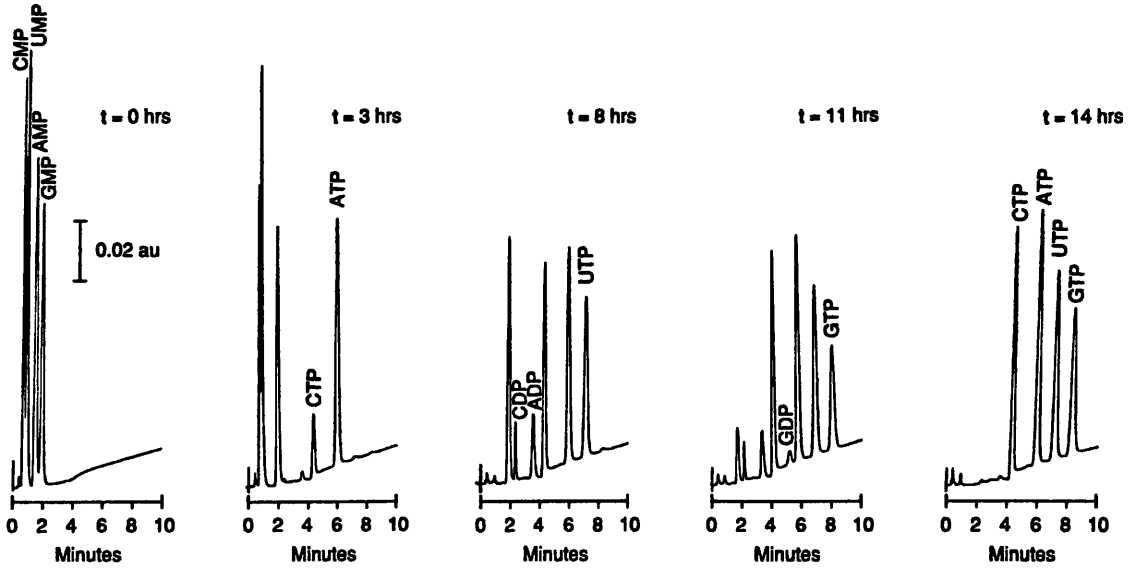


Figure A2.8: HPLC chromatograms of the reaction timecourse of an enzymatic nucleotide phosphorylation reaction. At the zero timepoint, a catalytic quantity of ATP was added to the reaction, along with myokinase, to initiate the reaction. After three hours, when the vast majority of AMP had been converted to ATP, guanylate kinase and nucleoside monophosphate kinase were added to initiate the phosphorylation of the other nucleoside monophosphates. The chromatograms are of 15 nmol of nucleotides injected onto a Vydac nucleotide analysis column using the elution scheme described in the text at a 2.0 ml/min flow rate and detecting at 268 mn at 0.2 aufs.

glycol (8000 molecular weight), 6.0 mM NTPs (~1.5 mM each), 200 nM each DNA strand, and T7 RNA polymerase (Grodberg & Dunn, 1988). Purification and preparation of the RNA for NMR was performed as described elsewhere (Wyatt *et al.*, 1991). The optimal added $MgCl_2$ concentration for transcription of the TAR template differed significantly for commercial NTPs (36 mM) and isotopically labeled cellular NTPs (9.6 mM $MgCl_2$). A comparison of reaction products from small scale transcription reactions for commercial and labeled NTPs is shown in Figure A2.9.

Practical Aspects

In order to generate large quantities of labeled nucleotide monomers, this procedure takes advantage of the abundance of RNA present in bacterial cells in log phase growth. It is estimated that 20% of the dry weight of rapidly dividing E. coli is RNA (Neidhardt, 1987), mainly in the form of ribosomal RNAs, and that 3% of the dry weight is chromosomal DNA. Labeled monomers can be prepared by isolation of total nucleic acids from bacterial cells grown on isotopic precursors, followed by conversion of the cellular RNA into a useable form. Here, the practical aspects of each step of this protocol are discussed.

Growth of E. coli

The growth of *E. coli* on glucose-minimal salts medium is straightforward (Miller, 1972; Sambrook *et al.*, 1989). The medium can be substituted with ^{15}N-ammonium salts and/or ^{13}C-glucose to produce uniformly labeled bacterial mass with high efficiency. For incorporation of ^{15}N only, *E. coli* is definitely the organism of choice. Optimization of a glucose-minimal salts medium for conversion of ^{15}N into biomass reduces the cost of the isotope to a small fraction of the total cost of preparing the nucleotides. However, for ^{13}C

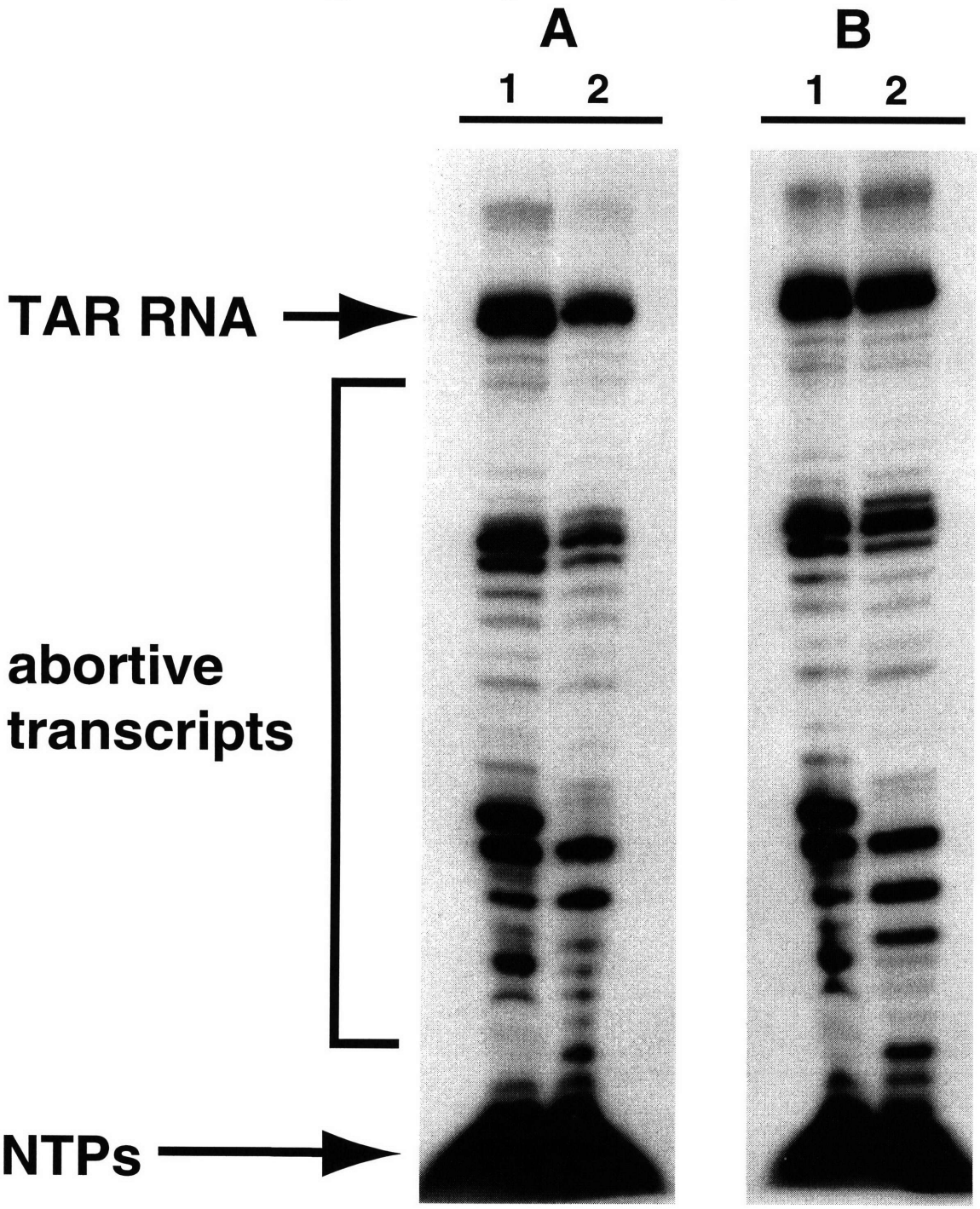


Figure A2.9: A comparison of *in vitro* transcription of TAR RNA performed with commercial nucleoside triphosphates and labeled NTPs. Small scale transcription reactions were performed as described, except that α-^{32}P labeled CTP was included in each reaction. (a) Transcription of TAR template in standard transcription buffer using Sigma NTPs (A1) and cellularly derived NTPs (A2). (b) Transcription of TAR template using Sigma NTPs in standard transcription buffer (B1) and cellularly derived NTPs in an optimized transcription buffer (B2). Lanes A1 and B1 differ slightly in the pattern of abortive transcripts compared to A2 and B2 due to the presence of GMP for priming transcription in the reactions containing commercial NTPs.

incorporation, the cost of the labeled glucose is quite high. On a per mole of carbon basis, ^{13}C-methanol is the most inexpensive isotopic source for carbon labeling (in the 1994-1995 Cambridge Isotope Laboratories catalog, ^{13}C-methanol is sold for $3,370 per mole and u-^{13}C-glucose is sold for $15,910 per mole), and thus, methylotrophic bacteria were exploited for the preparation of ^{13}C labeled ribonucleotides.

Growth of Methylophilus methylotrophus

There are two classes of methylotrophic bacteria that are distinguished by the pathway used to incorporate methanol carbons into cell material. The choice of methylotrophic strain for isotopic labeling requires consideration of the consequences of these two pathways. In the serine pathway (Anthony, 1982), the net reaction for assimilation of methanol is conversion of two moles of formaldehyde and one mole of CO_2 into one mole of 3-phosphoglycerate (Anthony, 1982). The net reaction in the ribulose monophosphate cycle (or, alternatively the hexulose monophosphate (HMP) pathway) is the conversion of three equivalents of methanol into 3-phosphoglycerate (Anthony, 1982). In contrast to the serine pathway of methanol assimilation, no carbon is incorporated into metabolic pathways from CO_2 via the HMP pathway. Thus, there is a significant danger of dilution of the input isotope by incorporation of atmospheric CO_2 in serine pathway organisms such as *Pseudomonas AM1* or *Methylobacterium extorquens*, but not in HMP pathway organisms. However, it has been reported that growth of *Methylobacterium extorquens* on 30%-^{13}C-methanol, did not give rise to any apparent isotopic dilution (Hines *et al.*, 1993). Nonetheless, the methylotrophic strain *Methylophilus methylotrophus* (ATCC 53528), which is an obligate methylotroph using the HMP pathway, was chosen for optimal incorporation of ^{13}C substrate.

Another primary consideration in isotopic labeling was rapid growth rate on methanol. *Methylophilus methylotrophus* reproducibly grows under the described

conditions with a doubling time of 1.9 hours, while the doubling rates of other methylotrophic bacteria can be considerably slower (Hines *et al.*, 1993). The percentage of RNA as dry mass of a cell growing in log phase is a function of the growth rate; the doubling time of the cell decreases, the percentage of cell dry weight that is RNA increases (Herbert, 1961). The relatively rapid doubling time of *Methylophilus methylotrophus* on methanol compared to other methylotrophic bacteria maximizes the amount of RNA yielded per input gram isotopic substrate.

The primary problem related to the growth of *Methylophilus methylotrophus* has been the production of a heteropolysaccharide sheath during logarithmic growth. Closely related methylotrophs are known to secrete heteropolysaccharides during logarithmic growth (Stirling, 1987). This growth characteristic generates bacteria that yield loosely packed, white pellets when centrifuged. These bacteria produced a great deal of non-nucleic acid material during isopropanol precipitations, and resulted in very low yields of cellular nucleic acids. This problem was initially solved by keeping the final cell density below 0.5 optical density units at 660 nm. Later, it was shown that during growth of this bacterium on methanol-minimal salts medium, the pH of the medium drops from 6.8 to below 6.2 at the end of the growth, due to the limited buffering capacity of the low phosphate buffer concentration in the medium. Continuing growth by addition of more methanol causes the pH to drop further, and we observe difficulties in cell harvest associated with heterosaccharide sheath production. This problem was solved by the addition of base during the growth in order to maintain the pH between 6.6 and 6.8.

The effects of other components of the medium was also examined. The growth of *M. methylotrophus* was optimized for the highest production of biomass per input gram of methanol by growing small trial cultures containing increasing amounts of methanol, as shown in Figure A2.10. Above 0.1% methanol, cell yields dropped dramatically, and therefore 0.1% was chosen as the optimal methanol concentration for high yield growth. These organisms are very sensitive to phosphate concentration, and it was found that low

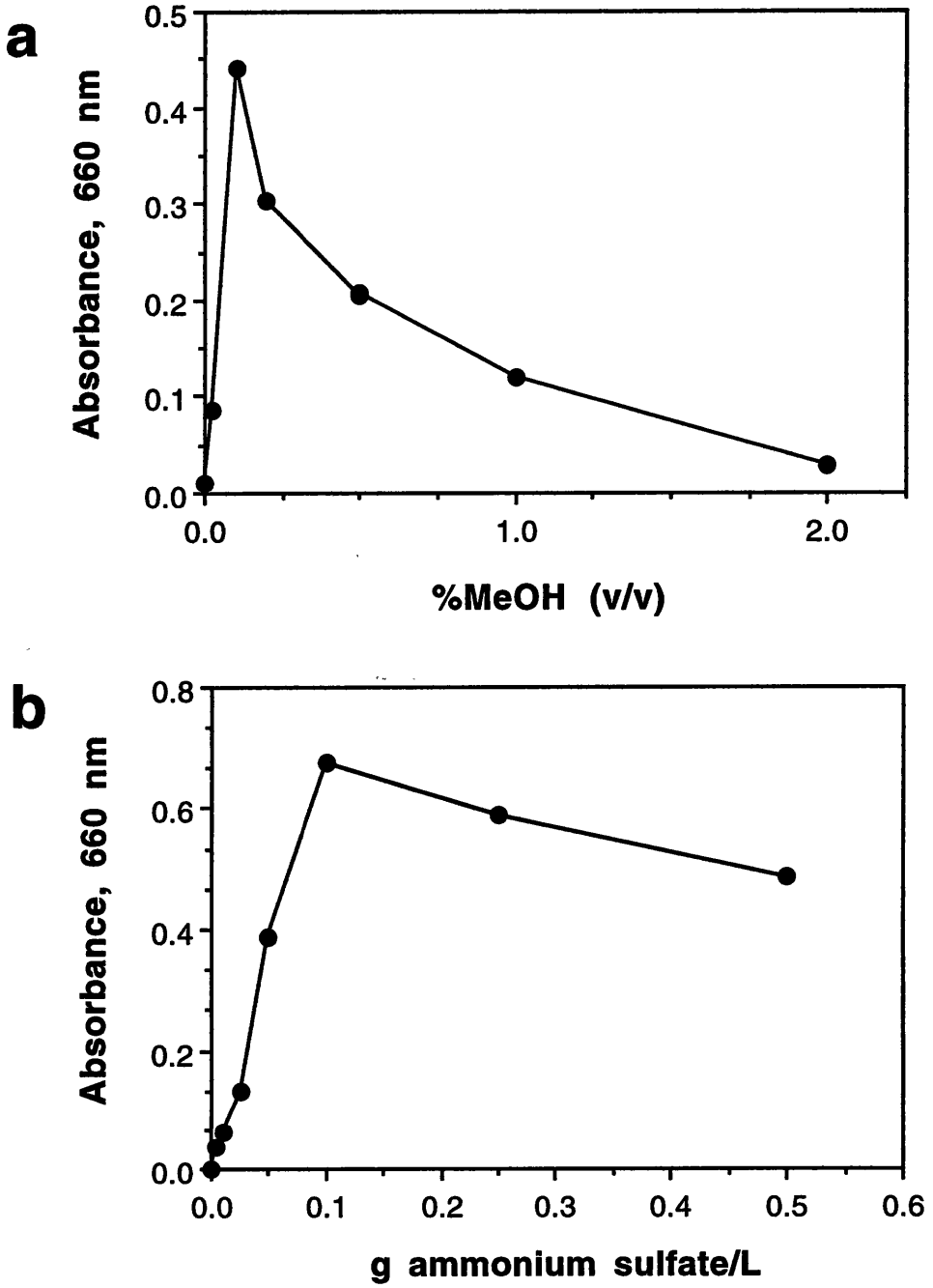


Figure A2.10: Yields of *M. methylotrophus* in batch culture as a function of (a) methanol and (b) nitrogen concentration. Cultures were grown to saturation (48 hrs) and the absorbance read at 660 nm.

phosphate media gave better growth rates (Anthony, 1982). Addition of a trace metal supplement containing EDTA as a chelating agent was found to inhibit growth. Neither an increase in the amount of trace metal supplement nor addition of vitamins (biotin, cyanocobalamin, thiamine, pantothenate, riboflavin, folic acid, and nicotinamide) to the medium had any effect on yields. A variety of growth media and conditions are reported in the literature (Long & Anthony, 1990) and we have observed a considerable range of growth rates under different conditions. The media reported here consistently gave the best results for this application. With optimization of the growth conditions, the yields of nucleic acids for *E. coli* and *Methylophilus methylotrophus* became comparable, as shown in Table A2.2.

Cellular Lysis

The highest yields of cellular nucleic acids were conveniently obtained by direct lysis of bacterial cells in hot phenol-chloroform solutions; elevated temperatures gave higher yields than room temperature lysis. Good yields of nucleic acids are correlated to minimizing the volume of the white proteinaceous interface layer obtained in the phenol extraction. Increasing the amount of cell mass per unit volume in the extraction reduces yields. Protease treatment greatly reduces the interfacial layer, and increases the yields of nucleic acids slightly, but we do not routinely use protease. Use of a Waring blender during the first phenol extraction reduced yields due to shearing of the nucleic acids. However, the blender treatment during the second and third phenol extractions increases the yield of these two extractions because it breaks up the inclusion layer efficiently. Isopropanol precipitation gives comparable yields to ethanol precipitation, but offers the advantage that only an equal volume of isopropanol is required for precipitation. Use of sodium acetate for the precipitation greatly increases the yield compared to sodium chloride.

Table A2.2: Yields of labeled ribonucleotides

Organism	Limiting nutrient	g cells / g isotope	mg NMPs / g cells	mg NMPs / g isotope
E. coli	$^{13}C_6$-glucose + $(^{15}NH_4)_2SO_4$	2.9	20	58
E. coli	$(^{15}NH_4)_2SO_4$	6.4	7	45
M. methylotrophus	^{13}C-CH_3OH[a]	2.5	10	25
M. methylotrophus	^{13}C-CH_3OH[b]	1.9	27	54
M. methylotrophus	^{13}C-CH_3OH + $(^{15}NH_4)_2SO_4$[b]	1.5	32	48

[a] *M. methylotrophus* grown in 1 L shake flasks (non-optimized conditions)
[b] *M. methylotrophus* grown in a 10 L fermentor (optimized conditions)

Nucleic acid hydrolysis

P1 nuclease rapidly degrades both DNA and RNA into 5' nucleoside monophosphates. DNA is digested much more slowly than RNA, and the hydrolysis of DNA is greatly enhanced by a denaturation step. We favor using large amounts of nuclease and a short digestion period since dephosphorylation of the nucleotides has been observed during prolonged P1 nuclease digestions. The digestion is readily monitored by reverse phase HPLC analysis of aliquots that are dephosphorylated prior to injection.

Nucleotide separation

We have avoided the use of DNase in the lysis procedure to permit isolation of the deoxyribonucleotides as a byproduct, and to avoid the expense of using large quantities of RNase free DNase. Boronate affinity chromatography rapidly and efficiently separated deoxyribonucleotides (dNMPs) from ribonucleotides (NMPs), and resulted in desalting and purification of the NMPs. Boronates form covalent complexes with 1,2 cis-diols at high pH that are readily hydrolyzed at low pH (Figure A2.11) (Mazzeo & Krull, 1989). There are several commercially available boronate affinity media based on *m*-aminophenylboronic acid attached to different polymeric resins. Of these, we have found the only satisfactory resin for this purpose to be Affi-gel 601, a polyacrylamide based resin. Other resins were unacceptable because of low nucleotide capacity or failure to bind NMPs. Affi-gel 601 does suffer from slow deterioration due to hydrolysis of the support, releasing *m*-aminophenylboronic acid. This compound gave a slight yellow color to the nucleotide preparations, but has no adverse effects on subsequent RNA synthesis.

Binding of NMPs requires high ionic strength, high pH (>9.0), and low temperatures, and dNMPs are not retained under these conditions. The dNMPs are collected in the flow-through as a byproduct, and the NMPs are eluted from the column using water that has been acidified with CO_2. Although the volume for elution is higher

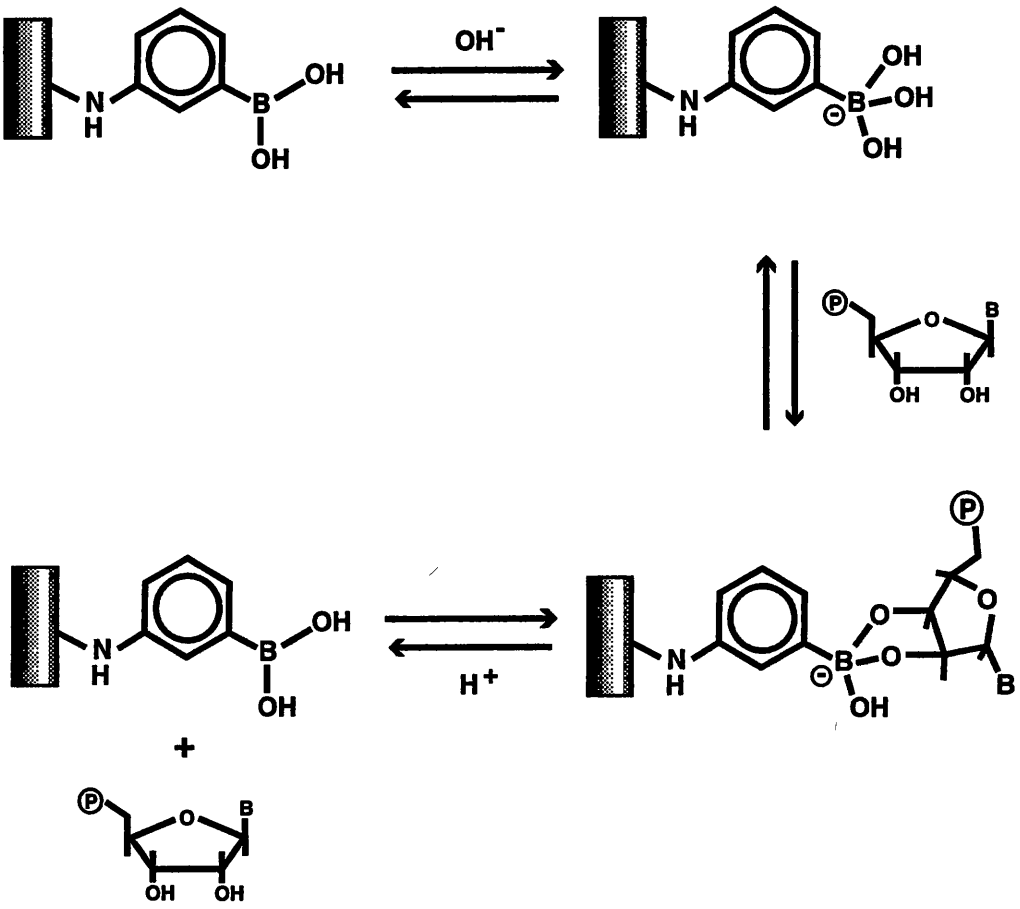


Figure A2.11: Interaction of an immobilized boronate ligand with the 1,2 cis-diol of the ribose sugar of a ribonucleoside monophosphate. Adapted from Mazzeo and Krull, 1989.

than if a buffer were used, a subsequent desalting step is avoided after concentration of the eluted NMPs. Affi-gel 601 undergoes large (~100%) volume changes upon change of pH and ionic strength, and use of a short wide column is recommended to avoid reduction of flow rate.

Separation of dNMPs from NMPs is readily monitored by reverse-phase HPLC of dephosphorylated aliquots of the column fractions. NMPs from this procedure are essentially free of dNMPs. The dNMP fraction contains a small (1-2%) amount of NMPs that leak through during the boronate chromatography. There is no evidence for detectable quantities of modified nucleotides that might be expected from tRNAs.

Separation of individual ribonucleotides

Ion exchange HPLC rapidly separates the four individual ribonucleoside monophosphates with linear scaleup capacity and baseline resolution. This column can resolve all four ribonucleotides with baseline resolution between >100 μg and 70 mg of total ribonucleotides without a significant loss in resolution. Reinjection of the pool of GMP onto the column reveals no cross-contamination with the other three nucleotides. HPLC separation of NMPs also has been successfully performed using a Nucleogen DEAE column to separate the individual nucleotides (Michnicka *et al.*, 1993). An alternative to HPLC is anion-exchange liquid chromatography using a AG1-X2 or AG1-X8 resin (Nikonowicz *et al.*, 1992; Hines *et al.*, 1993). The nucleotides are eluted from the column using either a linear gradient of increasing salt (Michnicka *et al.*, 1993) or step gradient of decreasing pH (Nikonowicz *et al.*, 1992).

The chemical and isotopic purity of the ribonucleotides were verified by ^{1}H NMR spectroscopy. Several milligrams of ^{13}C-labeled CMP was purified by HPLC and the ^{1}H NMR spectrum recorded. The H6 proton is split into a doublet by the large (~200 Hz) coupling to ^{13}C-C6. The isotopic purity of the nucleotides was estimated as 98% by

integration of the residual singlet from H6 coupled to a ^{12}C-C6 (Figure A2.12). The degree of deuteration of ribonucleotides derived from bacteria grown in D_2O was monitored in a similar fashion. The H6 resonance from the ^{1}H NMR spectrum of 52%-^{2}H rUMP is shown in Figure A2.13a, indicating approximately 50% deuteration at this position. The degree of deuteration at each position in the u-^{13}C-52% ^{2}H nucleotides was determined by integrating the multiplets that arise from ^{1}H coupling in the undecoupled ^{13}C NMR spectrum of u-^{13}C/52%-^{2}H rUMP, shown in Figure A2.13b. All of the carbons exhibited 52% deuteration at each site except for the ribose C1' position, which was only 39% deuterated, indicating nearly complete exchange of the methanol methyl protons with the solvent during growth.

Enzymatic phosphorylation

A general method for preparation of NTPs from cellular RNA has been reported by Whitesides (Haynie & Whitesides, 1990). We have employed this method in conversion of isotopically labeled NMPs to NTPs. A series of nucleotide kinases is used to complete the phosphorylation of NMPs to NDPs, and the final phosphorylation to NTP can be effected with either pyruvate kinase or acetate kinase. High concentrations of phosphoenol pyruvate (PEP) or acetyl phosphate (AcP) are used to drive the equilibrium toward NTPs.

Of the several possibilities for the phosphorylation of NDPs, we have settled on pyruvate kinase, where PEP is generated *in situ* from 3-phosphoglycerate using phosphoglycerate mutase and enolase, as illustrated in Figure A2.14. Both acetyl phosphate and PEP as ultimate phosphate donors give comparable results. We favor use of PEP primarily because a large quantity of precipitate formed during phosphorylations performed with acetate kinase, although phosphorylation still went virtually to completion. Since PEP is expensive compared to the enzymes and substrates required for *in situ* PEP generation, and both methods give comparable results, we favor the *in situ* method.

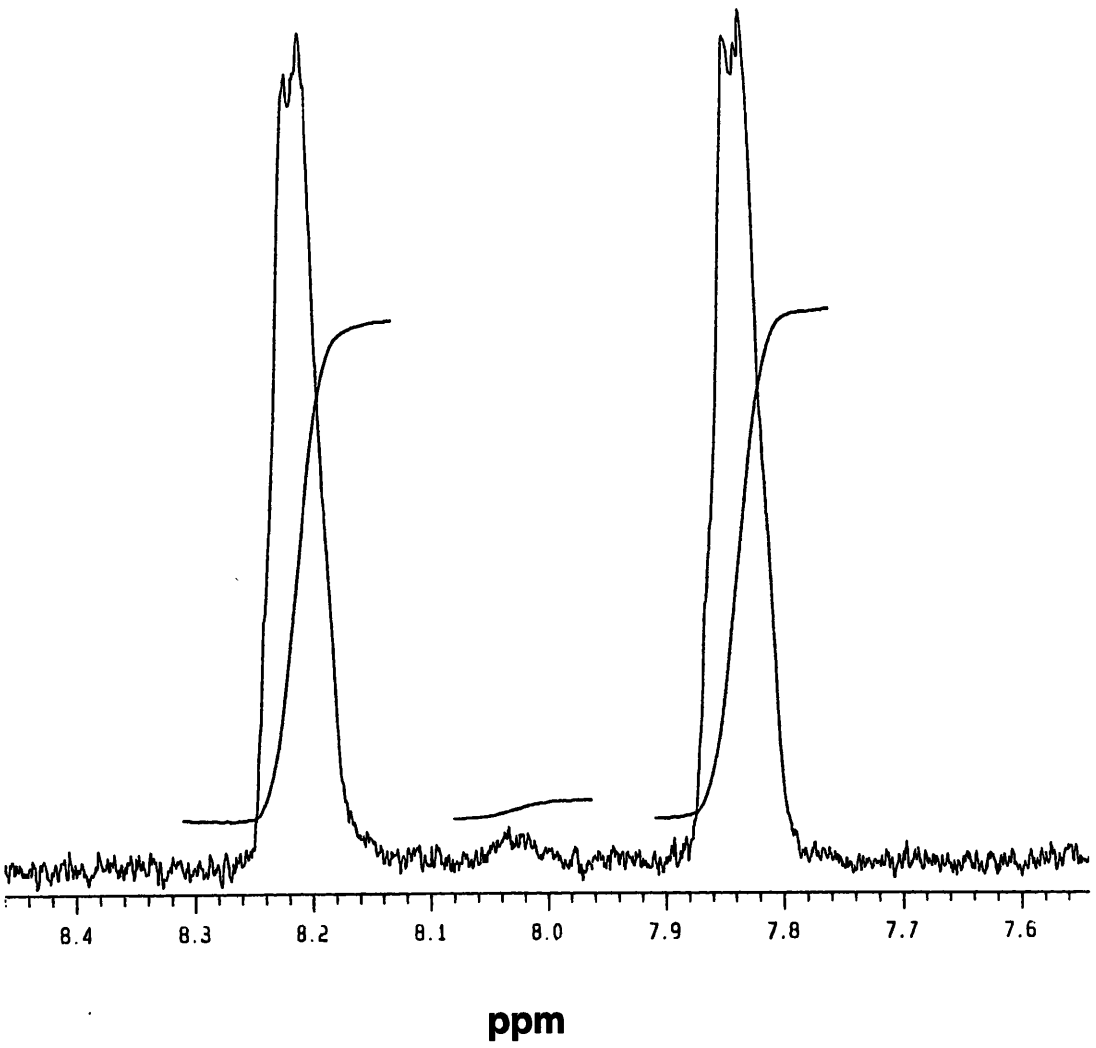


Figure A2.12: A portion of the ^{1}H NMR spectrum of u-^{13}C rCMP. The observed splitting pattern for the H6 proton arises from the superposition of a doublet due to spin-spin coupling to a ^{13}C at the C6 position, and a singlet arising from a H6 proton attached to a ^{12}C at the C6 position. Integration of the doublet and singlet demonstrates that this sample of cytosine is 98% ^{13}C-labeled.

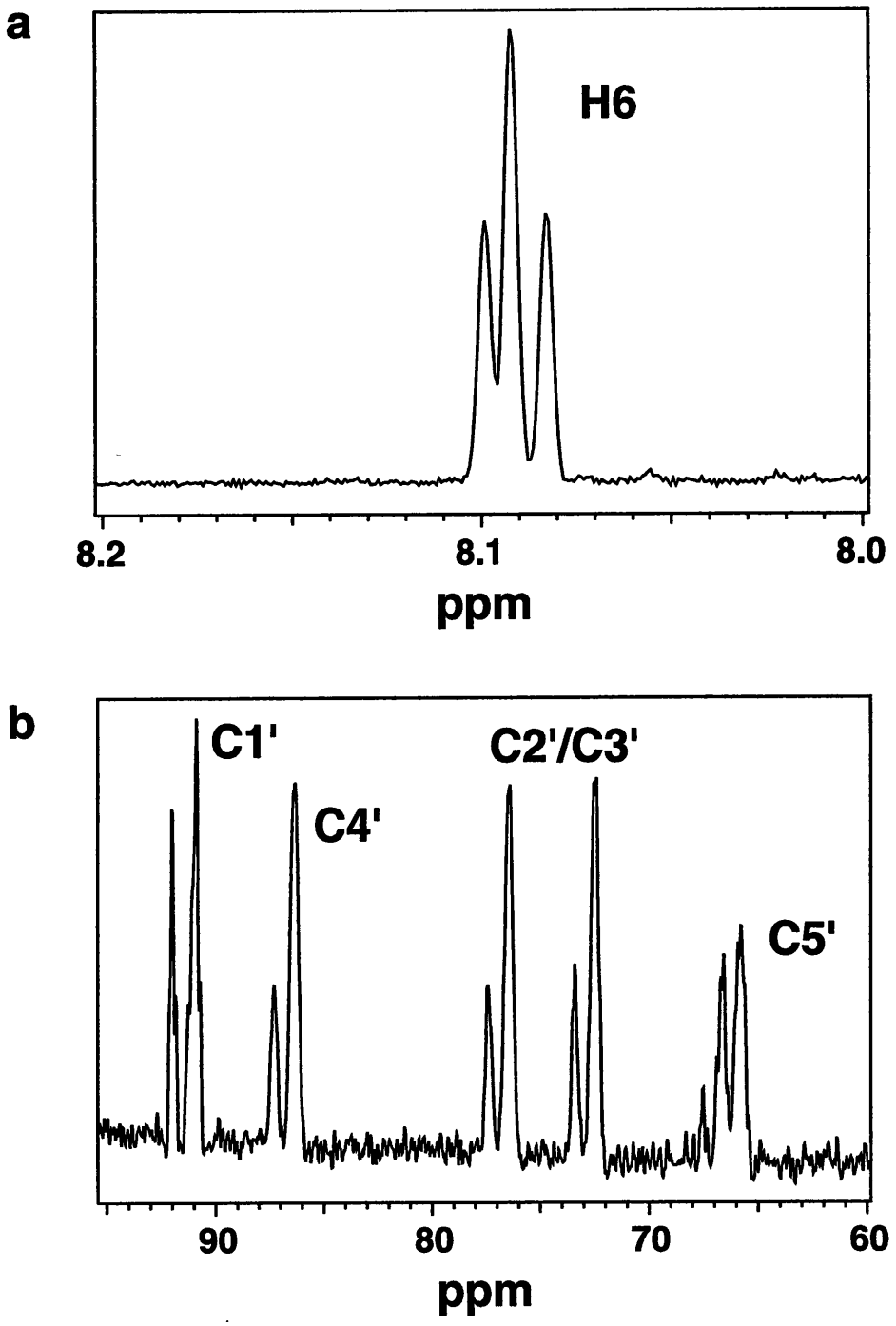


Figure A2.13: (a) A portion of the ^{1}H NMR spectrum of 52%-^{2}H rUMP. The observed pattern for the H6 resonance arises from superposition of a doublet due to spin-spin coupling to the proton on C5, and an isotope shifted singlet due to a deuteron on C5. (b) Ribose region of the ^{13}C NMR spectrum of u-^{13}C/52%-^{2}H rUMP. Each carbon resonance is a superposition of a doublet arising from carbons with coupling to the directly attached proton and an isotope shifted singlet arising from carbons bearing a deuteron.

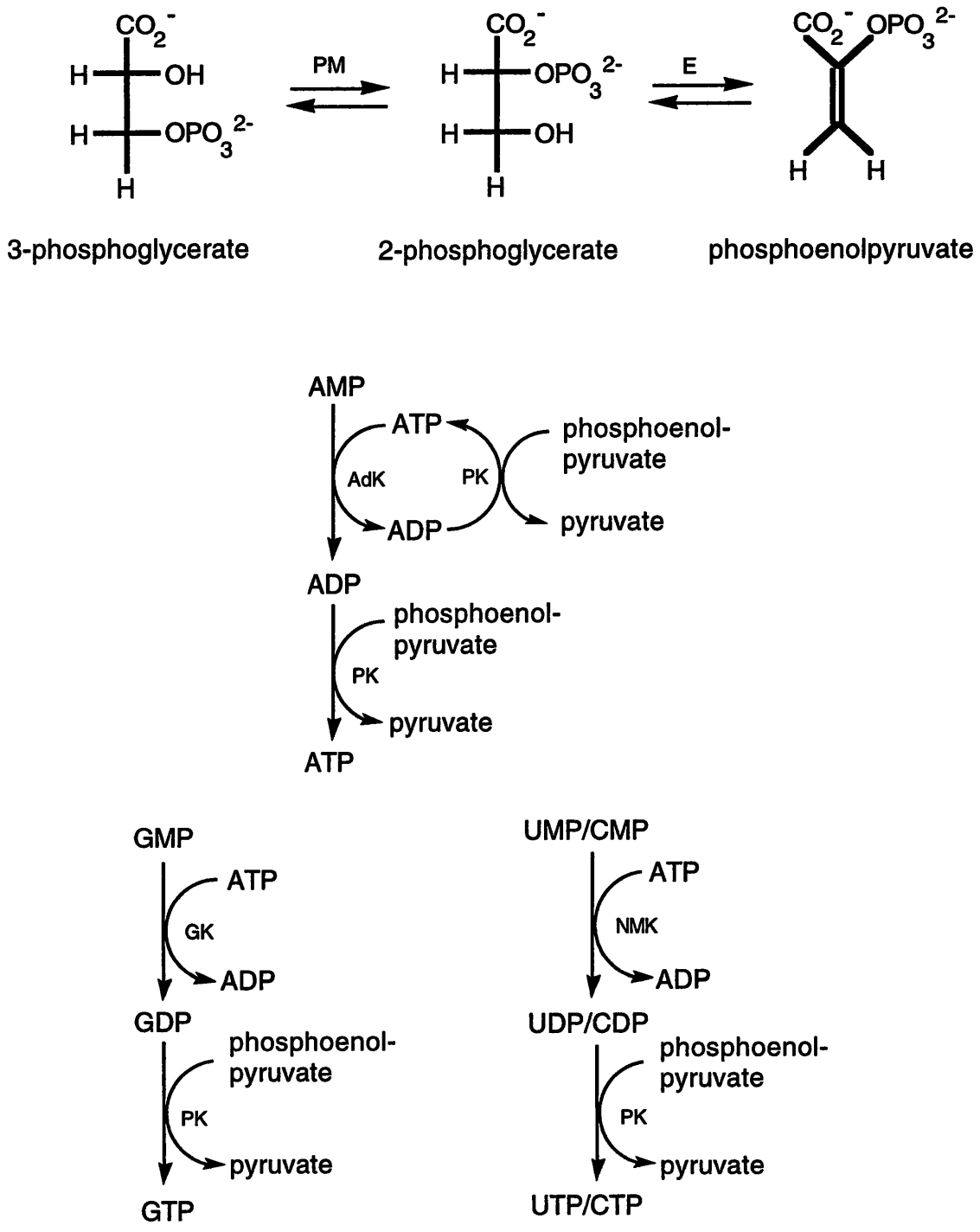


Figure A2.14: Phosphorylation of ribonucleoside monophosphates to triphosphates. Enzymes involved in this reaction are: (PM) phosphoglycerate mutase, (E) enolase, (AdK) adenylate kinase or myokinase, (PK) pyruvate kinase, (GK) guanylate kinase, and (NMK) nucleoside monophosphate kinase.

The enzymatic phosphorylation proceeds in two steps (Figure A2.14). First, ATP is generated from AMP using adenylate kinase and pyruvate kinase. The reaction is initiated by a catalytic amount (<0.1%) of unlabeled ATP. After several hours, ATP synthesis is complete, and the remaining nucleotide monophosphate kinases are added. The rate of phosphorylation of the remaining nucleotides is much more rapid when the total nucleotide concentration is kept at or below 10 mM. Some inhibition of the kinases may occur at high nucleotide concentrations. For the conditions we use, complete phosphorylation of all four NMPs to NTPs occurs in 14 hours. The enzymatic phosphorylation is monitored by ion exchange HPLC on a Vydac nucleotide analysis column.

Nucleotide purification

The NTPs from the enzymatic phosphorylation reaction are conveniently desalted and purified using boronate affinity chromatography. The binding of NTPs to the column is weaker than for NMPs, and consequently the elution volume for NTPs is usually smaller. NTPs do hydrolyze at pH 9.5 at an appreciable rate, and it is critical to chill high pH nucleotide solutions on ice at all times. If the total time at high pH is kept to a minimum (~2 hours), hydrolysis is <5%. NTPs taken directly from the phosphorylation reaction give no activity in transcription reactions. Desalted nucleotides exhibit increased transcription efficiency, but are still somewhat less efficient than commercial NTPs (*vide infra).* Finally, to ensure that the nucleotides are free from proteins, in particular RNases, the NTPs are ultrafiltered through a 10,000 molecular weight cutoff filter.

In Vitro Transcription

Isotopically labeled NTPs prepared as described were used in large scale *in vitro* transcription reactions using T7 RNA polymerase to prepare milligram quantities of RNAs

for NMR studies. A small scale transcription of TAR RNA is shown in Figure A2.9a using commercial NTPs (lane 1) and ^{13}C-labeled NTPs (lane 2), and the yield of full length transcript is 2-3 fold lower. Reoptimization of the Mg^{2+} concentration in the transcription reaction increased the efficiency of the labeled transcription reactions. Some Mg^{2+} may be complexed by the NTPs during the boronate affinity chromatography, and therefore, less Mg^{2+} is required in the transcription reactions than is typical. Comparison of the transcription efficiency of commercial and isotopic nucleotides under optimized conditions is shown in Figure A2.9b, and the yields are very similar. Also, consideration must be given to the yield of transcript per mole of input NTP. For the TAR template, the amount of transcripts increases only slightly as the labeled NTP concentrations are increased from 1 mM to 4 mM. Thus the optimal production of transcript per mole of NTP is obtained from lower NTP concentrations than are often used in unlabeled transcriptions.

NMR of u-^{13}C-labeled TAR

To demonstrate the utility of uniform isotopic labeling of RNA, we have performed a 3D-heteronuclear NMR experiment on u-^{13}C-labeled- TAR. Two problems encountered in NMR analysis of RNAs in particular are spectral crowding in the ribose proton region of the spectrum, and unobserved cross-peaks in homonuclear COSY experiments due to small coupling constants. Both of these problems are alleviated by the use of heteronuclear methods. An expansion of the H1'-H2' cross peak region of a DQF-COSY spectrum of TAR is shown in Figure A2.15. This region should contain one cross peak from each residue, but only 10 peaks are visible. This results from the small coupling constant between H1' and H2' in the C3'-*endo* sugar conformation usually observed in RNA. The only cross peaks usually observed in this region are from sugars in the C2'-*endo* conformation. The same region from two slices of different ^{13}C chemical shifts from a 3D-HCCH-COSY spectrum of u-^{13}C-TAR is shown in Figure A2.15b,c. In this experiment,

Figure A2.15: (next page): NMR spectra of TAR RNA. a) The region of a DQF-COSY spectrum of unlabeled TAR containing cross peaks between H1' and H2' protons. The magnitude spectrum was taken to permit direct comparison to the 3D spectra. b) The same regions as in (A) of a slice from a 3D-HCCH-COSY spectrum of u-^{13}C-TAR through ω_C=92.2 ppm. c) Slice of the same region through ω_C=90.0 ppm.

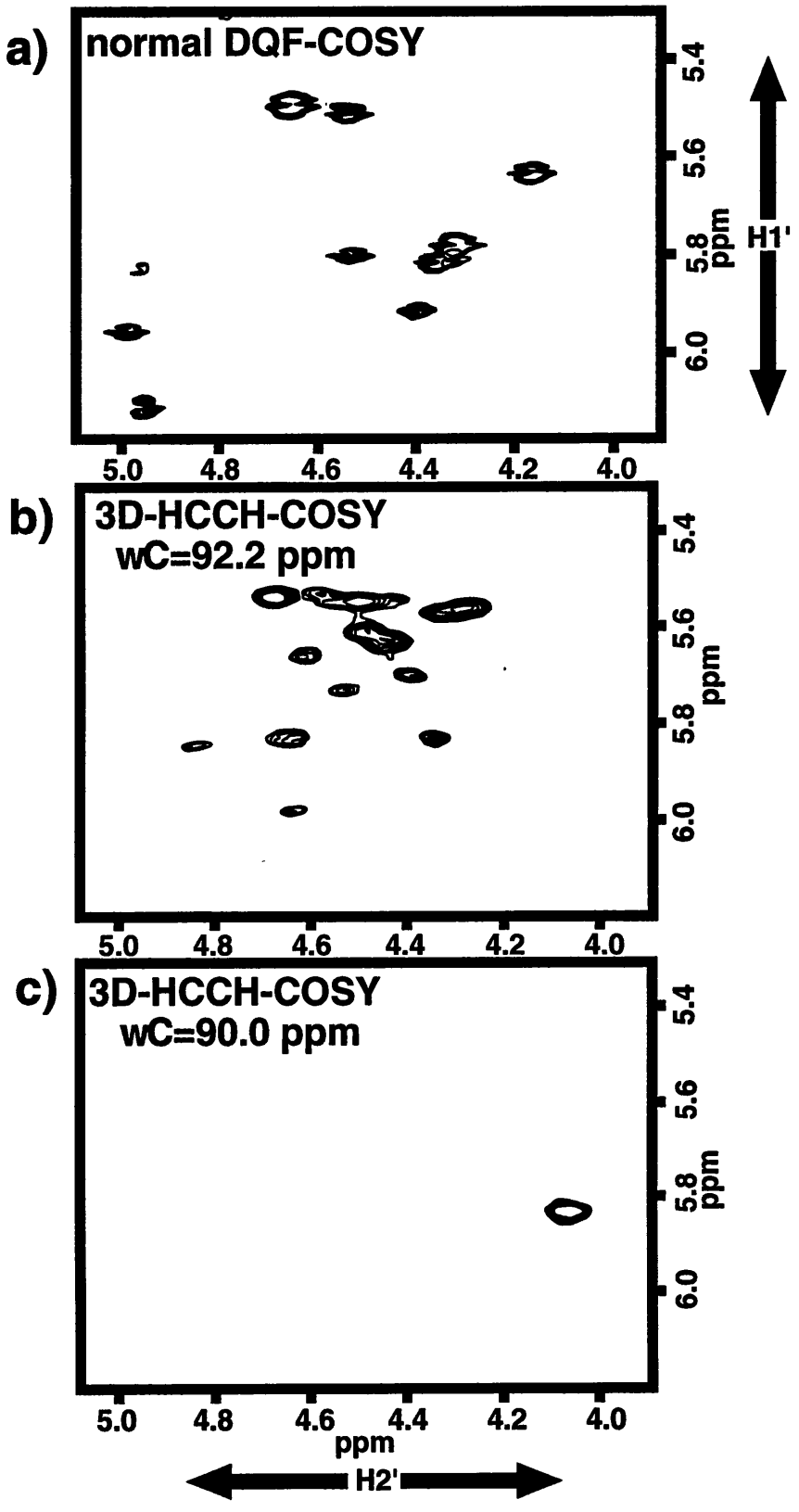

a)
normal DQF-COSY
5.4
5.6
5.8
6.0
ppm
H1'
5.0
4.8
4.6
4.4
4.2
4.0
b)
3D-HCCH-COSY
wC=92.2 ppm
5.4
5.6
5.8
6.0
ppm
5.0
4.8
4.6
4.4
4.2
4.0
c)
3D-HCCH-COSY
wC=90.0 ppm
5.4
5.6
5.8
6.0
ppm
5.0
4.8
4.6
4.4
4.2
4.0
ppm
H2'

Figure A2.15

the coherence transfer occurs through three large one-bond coupling constants that are independent of the sugar conformation, thus all 31 cross peaks can observed. In Figure A2.15b, many cross peaks are visible that are not present in Figure A2.15a. The dispersion of ^{13}C chemical shifts relieves spectral crowding in proton correlated spectra, as illustrated in Figure A2.15c, where only one cross peak is observed at that ^{13}C chemical shift.

Bibliography

Agrawal, R. K., Penczek, P., Grassucci, R. A., Li, Y., Leith, A., Nierhaus, K. H. & Frank, J. (1996). Direct visualization of A-, P-, and E-site transfer RNAs in the *Escherichia coli* ribosome. *Science* **271**, 1000-1002.

Allison, L. A., Romaniuk, P. J. & Bakken, A. H. (1991). RNA-protein interactions of stored 5S RNA with TFIIIA and ribosomal protein L5 during *Xenopus* oogenesis. *Dev. Biol.* **144**, 129-144.

Altman, S., Kirsebom, L. & Talbot, S. (1993). Recent studies of ribonuclease P. *FASEB J.* **7**, 7-14.

Amann, E., Brosius, J. & Ptashne, M. (1983). Vectors bearing a hybrid *trp-lac* promoter useful for regulated expression of cloned genes in *Escherichia coli*. *Gene* **5**, 167-178.

Anderson, A. C., Earp, B. E. & Frederick, C. A. (1996). Sequence variation as a strategy for crystallizing RNA motifs. *J. Mol. Biol.* **259**, 696-703.

Anthony, C. (1982). *The Biochemistry of Methylotrophs*, Academic Press, London.

Arndt, E., Scholzen, T., Kromer, W., Hatakeyama, T. & Kimura, M. (1991). Primary structures of ribosomal proteins from the archaebacterium *Halobacterium marismortui* and the eubacterium *Bacillus stearothermophilus*. *Biochimie* **73**, 657-668.

Baer, M. L., Houser, F., Loesch-Fries, L. S. & Gehrke, L. (1994). Specific RNA binding by amino-terminal peptides of alfalfa mosaic virus coat protein. *EMBO J.* **13**, 727-735.

Baron, C., Heider, J. & Bock, A. (1993). Interaction of translation factor SELB with the formate dehydrogenase H selenopolypeptide mRNA. *Proc. Natl. Acad. Sci. USA* **90**, 4181-4185.

Bartel, D. P. & Szostak, J. W. (1994). Study of RNA-protein recognition by *in vitro* selection. In *RNA-Protein Interactions* (Nagai, K. & Mattaj, I. W., eds.), pp. 248-268. Oxford University Press, Inc, New York.

Bartel, D. P., Zapp, M. L., Green, M. R. & Szostak, J. W. (1991). HIV-1 Rev regulation involves recognition of non-Watson-Crick base pair in viral RNA. *Cell* **67**, 529-536.

Baserga, S. J. & Steitz, J. A. (1993). The diverse world of small ribonucleoproteins. In *The RNA world* (Gesteland, R. & Atkins, J. F., eds.), pp. 359-381. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Bassi, G. S., Mollegaard, N.-E., Murchie, A. I. H., von Kitzing, E. & Lilley, D. M. J. (1995). Ionic interactions and the global conformations of the hammerhead ribozyme. *Struct. biol.* **2**, 45-55.

Bassi, G. S., Murchie, A. I. H. & Lilley, D. M. J. (1996). The ion-induced folding of the hammerhead ribozyme: Core sequence changes that perturb folding into the active conformation. *RNA* **2**, 756-768.

Batey, R. T., Battiste, J. L. & Williamson, J. R. (1995). Preparation of isotopically enriched RNAs for heteronuclear NMR. *Methods Enzymol.* **261**, 300-322.

Batey, R. T., Cloutier, N., Mao, H. & Williamson, J. R. (1996). Improved large scale culture of *Methylophilus methylotrophus* for $^{13}C/^{15}N$ labeling and random franctional deuteration of ribonucleotides. *Nucleic Acids Res.* **24**, 4836-4837.

Batey, R. T., Inada, M., Kujawinski, E., Puglisi, J. D. & Williamson, J. R. (1992). Preparation of isotopically ribonucleotides for multidimensional NMR spectroscopy of RNA. *Nucleic Acids Res.* **20**, 4515-4523.

Batey, R. T. & Williamson, J. R. (1996a). Interaction of the *Bacillus stearothermophilus* ribosomal protein S15 with 16 S rRNA: I. Defining the minimal RNA site. *J. Mol. Biol.* **261**, 536-549.

Batey, R. T. & Williamson, J. R. (1996b). Interaction of the *Bacillus stearothermophilus* ribosomal protein S15 with 16 S rRNA: II. Specificity determinants of RNA-protein recognition. *J. Mol. Biol.* **261**, 550-567.

Battiste, J. L., Mao, H., Rao, N. S., Tan, R., Muhandiram, D. R., Kay, L. E., Frankel, A. D. & Williamson, J. R. (1996). α helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. *Science* **273**, 1547-1551.

Battiste, J. L., Tan, R., Frankel, A. D. & Williamson, J. R. (1995). Assignment and modeling of the rev response element RNA bound to a rev peptide using ^{13}C-heteronuclear NMR. *J. Biol. NMR* **6**, 375-389.

Bax, A. & Grzesiek, S. (1993). Methodological Advances in Protein NMR. *Accounts of Chemical Research* **26**, 131-138.

Beardsmore, A. J., Aperghis, P. N. G. & Quayle, J. R. (1982). Characterization of the assimilatory and dissimilatory pathways of carbon metabolism during growth of *Methylophilus methylotrophus* in methanol. *J. Gen. Microb.* **128**, 1423-1439.

Benard, L., Philippe, C., Dondon, L., Grunberg-Manago, M., Ehresmann, B., Ehresmann, C. & Portier, C. (1994). Mutational analysis of pseudoknot structure of the S15 translational operator from *Escherichia coli. Mol. Microbiol.* **14**, 31-40.

Benevides, J. M., Lemeur, D. & Thomas, G. J. J. (1984). Molecular conformations and 8-CH exchange rates of purine ribo- and deoxyribonucleotides: Investigation by Raman spectroscopy. *Biopolymers* **23**, 1011-1024.

Berglund, H., Rak, A., Garber, M. & Härd, T. (1996). *XVIIth International Conference on Magnetic Resonance in Biological Systems, Keystone, Colorado.*

Biou, V., Yaremchuk, A., Tukalo, M. & Cusack, S. (1994). The 2.9 Å crystal structure of *T. thermophilus* seryl-tRNA synthetase complexed with $tRNA^{Ser}$. *Science* **263**, 1404-1410.

Blackburn, E. H. (1993). Telomerase. In *The RNA world* (Gesteland, R. & Atkins, J. F., eds.), pp. 557-576. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Brimacombe, R. (1995). The structure of ribosomal RNA: a three-dimensional jigsaw puzzle. *Eur. J. Biochem.* **230**, 365-383.

Brimacombe, R., Atmadja, J., Stiege, W. & Schuler, D. (1988). A detailed model of the three-dimensional structure of *Escherichia coli* 16 S ribosomal RNA *in situ* in the 30 S subunit. *J. Mol. Biol.* **199**, 115-136.

Brookes, P. & Lawley, P. D. (1961). The reaction of mono- and di-functional alkylating agents with nucleic acids. *Biochem. J.* **80**, 496-503.

Brosius, J. & Holy, A. (1984). Regulation of ribosomal RNA promoters with a synthetic *lac* operator. *Proc. Natl. Acad. Sci. USA* **81**, 6929-6933.

Brown, B., Bowie, J. U. & Sauer, R. T. (1990). Arc repressor is tetrameric when bound to operator DNA. *Biochemistry* **29**, 11189-11195.

Brunel, C., Romby, P., Westhof, E., Ehresmann, C. & Ehresmann, B. (1991). Three-dimensional model of *Escherichia coli* ribosomal 5 S RNA as deduced from structure probing in solution and computer modeling. *J. Mol. Biol.* **221**, 293-308.

Brush, C. K., Stone, M. P. & Harris, T. M. (1988). Selective reversible deuteration of oligodeoxynucleotides: Simplification of two-dimensional nuclear Overhauser effect NMR spectral assignment of a non-self-complementary dodecamer duplex. *Biochemistry* **27**, 115-122.

Calnan, B. J., Tidor, B., Biancalana, S., Hudson, D. & Frankel, A. D. (1991). Arginine-mediated RNA recognition: The arginine fork. *Science* **252**, 1167-1171.

Capel, M. S., Engelman, D. M., Freeborn, B. R., Kjeldgaard, M., Langer, J. A., Ramakrishnan, V., Schindler, D. P., Schneider, D. K., Schoenborn, B. P., Sillers, I. Y., Yabuki, S. & Moore, P. B. (1987). A complete mapping of the ribosomal proteins in the small ribosomal subunit of *Escherichia coli*. *Science* **238**, 1403-1406.

Caponigro, G. & Parker, R. (1996). Mechanisms and control of mRNA turnover in *Saccharomyces cerevisiae*. *Microbiol. Rev.* **60**, 233-249.

Caprara, M. G., Mohr, G. & Lambowitz, A. M. (1996). A tyrosyl-tRNA synthetase protein induces tertiary folding of the group I intron catalytic core. *J. Mol. Biol.* **257**, 512-531.

Carey, J. (1988). Gel retardation at low pH resolves *trp* repressor-DNA complexes for quantitative study. *Proc. Natl. Acad. Sci. USA* **85**, 975-979.

Carey, J. (1991). Gel retardation. *Methods Enzymol.* **208**, 103-117.

Cashmore, A. R. & Peterson, G. B. (1978). The degradation of DNA by hydrazine: identification of 3-ureidopyrazole as a product of the hydrzinolysis of deoxycytidylic acid residues. *Nucleic Acids Res.* **5**, 2485-2491.

Cate, J. H. & Doudna, J. A. (1996). Metal-binding sites in the major groove of a large ribozyme domain. *Structure* **4**, 1221-1229.

Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R. & Doudna, J. A. (1996). Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* **273**, 1678-1685.

Cerretti, D. P., Mettheakis, L. C., Kearney, K. R., Vu, L. & Nomura, M. (1988). Translational regulation of the *spc* operon in *Escherichia coli*: Identification and structural analysis of the target site for S8 repressor protein. *J. Mol. Biol.* **204**, 309-329.

Chastain, M. & Tinoco, I., Jr. (1991). Structural elements in RNA. *Prog. Nucleic Acids Res. Mol. Biol.* **41**, 131-177.

Cheong, C., Varani, G. & Tinoco, I., Jr. (1990). Solution structure of an unusually stable RNA hairpin, 5'GGAC(UUCG)GUCC. *Nature* **346**, 680-682.

Clackson, T., Gussow, D. & Jones, P. T. (1991). General applications of PCR to gene cloning and manipulation. In *PCR: A practical approach* (McPherson, M. J., Quirke, P. & Taylor, G. R., eds.), pp. 187-214. Oxford University Press, Oxford.

Clegg, R. M., Murchie, A. I. H. & Lilley, D. M. J. (1994). Solution structure of the four-way DNA junction at low salt conditions: a fluorescence resonance energy transfer analysis. *Biophysical Journal* **66**, 99-109.

Clore, G. M. & Gronenborn, A. M. (1991). Applications of three- and four-dimensional heteronuclear NMR spectroscopy to protein structure determination. *Prog. Nucl. Magn. Reson. Spectroc.* **23**, 43-92.

Clore, G. M., Kay, L. E., Bax, A. & Gronenborn, A. (1991). Four-dimensional $^{13}C/^{13}C$-edited nuclear Overhauser enhancement spectroscopy of a protein in solution: application to interleukin 1β. *Biochemistry* **30**, 12-18.

Cohen, S. N. (1995). Surprises at the 3' end of prokaryotic RNA. *Cell* **80**, 829-832.

Conway, L. & Wickens, M. (1987). Analysis of mRNA 3' end formation by modificatin interference: the only modifications which prevent processing lie in AAUAAA and the poly(A) site. *EMBO J.* **6**, 4177-4184.

Conway, L. & Wickens, M. (1989). Modification interference analysis of reactions using RNA substrates. In *Methods in Enzymology* (Dahlberg, J. E. & Abelson, J. N., eds.), Vol. 180, pp. 369-379. Academic Press, Inc., San Diego.

Cooper, J. P. & Hagerman, P. J. (1987). Gel electrophoretic analysis of the geometry of a DNA four-way junction. *J. Mol. Biol.* **198**, 711-719.

Cooper, J. P. & Hagerman, P. J. (1989). Geometry of a branched DNA structure in solution. *Proc. Natl. Acad. Sci. USA* **86**, 7336-7340.

Cooperman, B. S., Weitzman, C. J. & Buck, M. A. (1988). Reversed-Phase High-Performance liquid chromatography of ribosomal proteins. In *Methods in Enzymology*, Vol. 164, pp. 523-532.

Cotton, F. A., Wilkinson, G. & Gaus, P. L. (1987). *Basic Inorganic Chemistry*, John Wiley & Sons, New York.

Creighton, T. E. (1984). *Proteins: Structures and molecular principles*, W. H. Freeman and Company, New York.

Cullen, B. R. & Malim, M. H. (1991). The HIV-1 Rev protein: prototype of a novel class of eukaryotic post-transcriptional regulators. *Trends Biochem. Sci.* **16**, 346-350.

Curcio, M. J. & Belfort, M. (1996). Retrohoming: cDNA-mediated mobility of group II introns requires a catalytic RNA. *Cell* **84**, 9-12.

Dahlberg, A. E., Dingman, C. W. & Peacock, A. C. (1969). Electrophoretic characterization of bacterial polyribosomes in agarose-acrylamide composite gels. *J. Mol. Biol.* **41**, 139-147.

Davanloo, P., Rosenberg, A. H., Dunn, J. J. & Studier, F. W. (1984). Cloning and expression of the gene for bacteriophage T7 RNA polymerase. *Proc. Natl. Acad. Sci. USA* **81**, 2035-2039.

Davies, C., Ramakrishnan, V. & White, S. W. (1996). Structural evidence for specific S8-RNA and S8-protein interactions within the 30S ribosomal subunit: ribosomal protein S8 from *Bacillus stearothermophilus* at 1.9 Å resolution. *Structure* **4**, 1093-1104.

de Boer, H. A. & Kastelein, R. A. (1986). Biased codon usage: an exploration of its role in optimization of translation. In *Maximizing Gene Expression* (Reznikoff, W. & Gold, L., eds.), pp. 225-255. Butterworths, Boston.

Deckman, I. C. & Draper, D. E. (1985). Specific interaction between ribosomal protein S4 and the α operon messenger RNA. *Biochemistry* **24**, 7860-7865.

Derome, A. E. (1987). *Modern NMR Techniques for Chemistry Research*. Organic Chemistry Series (Baldwin, J. E., Ed.), 6, Pergamon Press, Oxford.

Dervan, P. B. (1986). Design of sequence-specific DNA-binding molecules. *Science* **232**, 464-471.

Dieckmann, T., Suzuki, E., Nakamura, G. K. & Feigon, J. (1996). Solution structure of an ATP-binding RNA aptamer reveals a novel fold. *RNA* **2**, 628-640.

Donis-Keller, H., Maxam, A. & Gilbert, W. (1977). Mapping adenines, guanines, and pyrimidines in RNA. *Nucleic Acids Res.* **4**, 2527-2538.

Doudna, J. A., Grosshans, C., Gooding, A. & Kundrot, C. E. (1993). Crystallization of ribozymes and small RNA motifs by a sparse matrix approach. *Proc. Natl. Acad. Sci. USA* **90**, 7829-7833.

Dragon, F. & Brakier-Gingras, L. (1993). Interaction of *Escherichia coli* ribosomal protein S7 with 16S rRNA. *Nucleic Acids Res.* **21**, 1199-1203.

Dragon, F., Payant, C. & Brakier-Gingras, L. (1994). Mutational and structural analysis of the RNA binding site for *Escherichia coli* ribosomal protein S7. *J. Mol. Biol.* **244**, 74-85.

Dreyfuss, G., Hentze, M. & Lamond, A. I. (1996). From transcript to protein. *Cell* **85**, 963-972.

Duckett, D. R. & Lilley, D. M. J. (1990). The three-way DNA junction is a Y-shaped molecule in which there is no helix-helix stacking. *EMBO J.* **9**, 1659-1664.

Duckett, D. R. & Lilley, D. M. J. (1991). Effects of base mismatches on the structure of the four-way DNA junction. *J. Mol. Biol.* **221**, 147-161.

Duckett, D. R., Murchie, A. I. H., Diekmann, S., Von Kitzing, E., Kemper, B. & Lilley, D. M. J. (1988). The structure of the Holliday junction and its resolution. *Cell* **55**, 79-89.

Duckett, D. R., Murchie, A. I. H. & Lilley, D. M. (1995). The global folding of four-way helical junctions in RNA, including that in U1 snRNA. *Cell* **83**, 1027-1036.

Ehrenberg, L., Fedorcsak, I. & Solymosy, F. (1976). Diethyl pyrocarbonate in nucleic acid research. In *Prog. Nucl. Acid Res. Mol. Biol.* (Cohen, W. E., ed.), Vol. 16, pp. 189-262. Academic Press, Inc., New York.

Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J.-P. & Ehresmann, B. (1987). Probing the structure of RNAs in solution. *Nucleic Acids Res.* **15**, 53-72.

England, T. E. & Uhlenbeck, O. C. (1978). 3'-Terminal labelling of RNA with T4 RNA ligase. *Nature* **275**, 560-561.

Fan, P., Suri, A. K., Fiala, R., Live, D. & Patel, D. J. (1996). Molecular recognition in the FMN-RNA aptamer complex. *J. Mol. Biol.* **258**, 480-500.

Farmer, B. T., II, Müller, L., Nikonowicz, E. P. & Pardi, A. (1993). Unambiguous resonance assignments in carbon-13, nitrogen-15-labeled nucleic acids by 3D triple-resonance NMR. *J. Am. Chem. Soc.* **115**, 11040-1.

Farmer, B. T., II, Müller, L., Nikonowicz, E. P. & Pardi, A. (1994). Unambiguous through-bond sugar-to-base correlations for purines in ^{13}C, ^{15}N-labeled nucleic acids: the $H_sC_sN_b$, $H_sC_s(N)_bC_b$, and $H_bN_bC_b$ experiments. *J. Biol. NMR* **4**, 129-133.

Ferré-D'Amaré, A. R. & Doudna, J. A. (1996). Use of *cis*- and *trans*-ribozymes to remove 5' and 3' heterogeneities from milligrams of in vitro transcribed RNA. *Nucleic Acids Res.* **24**, 977-978.

Fersht, A. (1985). *Enzyme Structure and Mechanism*, W. H. Freeman and Company, New York.

Földesi, A., Nilson, F. P. R., Glemarec, C., Gioeli, C. & Chattopadhyaya, J. (1992). Synthesis of 1'#,2',3',4'#,5',5''-$^{2}H_{6}$-β-D- ribonucleosides and 1'#,2',2'',3',4'#,5',5''-$^{2}H_{7}$-β-D-2'- deoxyribonucleosides for selective suppression of proton resonances in partially- deuterated oligo-DNA , oligo-RNA and in 2,5A core (^{1}H NMR window). *Tetrahedron* **48**, 9033-9072.

Földesi, A., Yamakage, S.-I., Nilsson, F. P. R., Maltseva, T. V. & Chattopadhyaya, J. (1996). The use of non-uniform deuterium labelling ['NMR-window'] to study the NMR structure of a 21mer RNA hairpin. *Nucleic Acids Res.* **24**, 1187-1194.

Frankel, A. D. (1992). Activation of HIV transcription by Tat. *Curr. Opin. Genes Dev.* **2**, 293-298.

Fried, M. & Crothers, D. M. (1981). Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res.* **9**, 6505-6525.

Fried, M. G. & Crothers, D. M. (1984). Kinetics and mechanism in the reaction of gene regulatory proteins in DNA. *J. Mol. Biol.* **172**, 263-282.

Fritz, C. C. & Green, M. R. (1996). HIV Rev uses a conserved cellular protein export pathway for the nucleocytoplasmic transport of viral RNAs. *Curr. Biol.* **6**, 848-854.

Garner, M. M. & Revzin, A. (1981). A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory system. *Nucleic Acids Res.* **9**, 3047-3060.

Garrett, R. A., Rak, K. H., Daya, L. & Stoffler, G. (1971). Ribosomal Proteins. XXIX. Specific protein binding sites on 16S rRNA of *Escherichia coli*. *Mol. Gen. Genet.* **114**, 112-124.

Gaur, R. K. & Krupp, G. (1993). Modification interference approach to detect ribose moieties important for the optimal activity of a ribozyme. *Nucleic Acids Res.* **21**, 21-26.

Gautheret, D., Konings, D. & Gutell, R. R. (1995). G-U base pairing motifs in ribosomal RNA. *RNA* **1**, 807-814.

Gibbs, C. S. & Zoller, M. J. (1991). Identification of functional residues in proteins by charged-to-alanine scanning mutagenesis. *Methods* **3**, 165-173.

Gilbert, W. (1986). The RNA world. *Nature* **319**, 618.

Gish, G. & Eckstein, F. (1988). DNA and RNA sequence determination base on phophorthioate chemistry. *Science* **240**, 1520-1522.

Glotz, C., Mussig, J., Gewitz, H.-S., Jakowski, I., Arad, T., Yonath, A. & Wittman, H. G. (1987). Three-dimensional crystals of ribosomes and their subunits from eu- and archeabacteria. *Biochem. Int.* **15**, 953-960.

Gogia, Z. V., Venyaminov, S. Y., Bushuev, V. N., Serdyuk, I. N., Lim, V. I. & Spirin, A. S. (1979). Compact globular structure of protein S15 from *Escherichia coli* ribosomes. *FEBS Lett.* **105**, 63-69.

Greenbaum, N. L. (1996). How Tat targets TAR: structure of the BIV peptide-RNA complex. *Structure* **4**, 5-9.

Greenblatt, J., Nodwell, J. R. & Mason, S. W. (1993). Transcriptional antitermination. *Nature* **364**, 401-406.

Gregory, R. R., Cahill, P. B. F., Thurlow, D. L. & Zimmerman, R. A. (1988). Interaction of *Escherichia coli* ribosmal protein S8 with its binding sites in ribosomal RNA and messenger RNA. *J. Mol. Biol.* **204**, 295-307.

Grodberg, J. & Dunn, J. J. (1988). OmpT encodes the *Escherichia coli* outer-membrane protease that cleaves T7-RNA polymerase during purification. *J. Bact.* **170**, 1245-1253.

Gubser, C. C. & Varani, G. (1996). Structure of the polyadenylation regulatory element of the human U1A pre-mRNA 3'-untranslated region and interaction with the U1A protein. *Biochemistry* **35**, 2253-2267.

Guerrier-Takada, C. & Altman, S. (1984). Catalytic activity of an RNA molecule prepared by transcription *in vitro*. *Science* **223**, 285-286.

Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849-857.

Gutell, R. R. (1994). Collection of small subunit (16S- and 16S-like) ribosomal RNA structures: 1994. *Nucleic Acids Res.* **22**, 3502-3507.

Gutell, R. R., Larsen, N. & Woese, C. (1994). Lessons from an evolving RNA: 16S and 23S rRNA structures from a comparative perspective. *Microbiol. Rev.* **58**, 10-26.

Gutell, R. R., Weiser, B., Woese, C. R. & Noller, H. F. (1985). Comparative anatomy of 16-S-like ribosomal RNA. *Prog. Nucleic Acid Res. Mol. Biol.* **32**, 155-216.

Gutell, R. R. & Woese, C. R. (1990). Higher-order structural elements in ribosomal RNAs: pseudoknots and the use of noncanonical pairs. *Proc. Natl. Acad. Sci. USA* **87**, 663-667.

Guthrie, C. (1991). Messenger RNA splicing in yeast: clues to why the spliceosome is a ribonucleoprotein. *Science* **253**, 157-163.

Hall, K. B. & Stump, W. T. (1992). Interaction of N-terminal domain of U1A protein with an RNA stem/loop. *Nucleic Acids Res.* **20**, 4283-4290.

Hamy, F., Asseline, U., Grasby, J., Shigenori, I., Pritchard, C., Slim, G., Butler, J. G., Karn, J. & Gait, M. J. (1993). Hydrogen-bonding contacts in the major groove are

required for Human Immunodeficiency Virus type-1 *tat* protein recognition of TAR RNA. *J. Mol. Biol.* **230**, 111-123.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166**, 557-580.

Hayatsu, H. (1976). Bisulfite modification of nucleic acids and their constituents. *Prog. Nucleic Acid Res. Mol. Biol.* **16**, 75-124.

Haynie, S. L. & Whitesides, G. M. (1990). Preparation of a mixture of nucleoside triphosphates suitable for use in synthesis of nucleotide phosphate sugars from ribonucleic acid using nuclease P1, a mixture of nucleoside monophosphokinases and acetate kinase. *Appl. Biochem. Biotech.* **23**, 205-220.

Held, W. A., Ballou, B., Mizushima, S. & Nomura, M. (1974). Assembly mapping of 30 S ribosomal proteins from *Escherichia coli*. Further studies. *J. Biol. Chem.* **249**, 3103-3111.

Held, W. A., Mizushima, S. & Nomura, M. (1973). Reconstitution of *Escherichia coli* 30 S ribosomal subunits from purified molecular components. *J. Biol. Chem.* **248**, 5720-5730.

Henderson, E., Hardin, C. C., Walk, S. K., Tinoco, I., Jr. & Blackburn, E. (1987). Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs. *Cell* **51**, 899-908.

Herbert, D. (1961). The chemical composition of micro-organisms as a function of their environment. In *Microbial reaction to environment: Eleventh symposium of the Society for General Microbiology*, Vol. 11, pp. 391-416. University Press, Cambridge.

Herr, W., Chapman, N. M. & Noller, H. F. (1979). Mechanism of ribosomal subunit association: Discrimination of specific sites in 16 S RNA essential for association activity. *J. Mol. Biol.* **130**, 433-439.

Heus, H. & Pardi, A. (1991). Structural features that give rise to unusual stability of RNA hairpins containing GNRA loops. *Science* **253**, 191-194.

Heus, H. A., Wijmenga, S. S., van de Ven, F. J. M. & Hilbers, C. W. (1994). Sequential backbone assignment in ^{13}C-labeled RNA via through-bond coherence transfer using three-dimensional triple resonance spectroscopy (^{1}H, ^{13}C, ^{31}P) and two-dimensional hetero TOCSY. *J. Am. Chem. Soc.* **116**, 4983-4984.

Higo, K., Held, W., Kahan, L. & Nomura, M. (1973). Functional correspondence between 30S ribosomal proteins of *Escherichia coli* and *Bacillus stearothermophilus*. *Proc. Natl. Acad. Sci. USA* **70**, 944-948.

Hines, J. V., Varani, G., Landry, S. M. & Tinoco, I., Jr. (1993). The stereospecific assignment of H5' and H5" in RNA using the sign of two-bond carbon-proton scalar couplings. *J. Am. Chem. Soc.* **115**, 11002-3.

Hore, P. J. (1983). Solvent suppression in fourier transform Nuclear Magnetic Resonance. *J. Mag. Res.* **55**, 283-300.

Hou, Y.-M. & Schimmel, P. (1988). A simple structural feature is a major determinant of the identity of a transfer RNA. *Nature* **333**, 140-145.

Howe, P. W. A., Nagai, K., Neuhaus, D. & Varani, G. (1994). NMR studies of U1 snRNA recognition by the N-terminal RNP domain of the human U1A protein. *EMBO J.* **13**, 3873-3881.

Hubbard, J. M. & Hearst, J. E. (1991). Computer modeling 16S ribosomal RNA. *J. Mol. Biol.* **221**, 889-907.

Huber, P. W. & Wool, I. G. (1984). Nuclease protection analysis of ribonucleoprotein complexes: use of the cytotoxic ribonuclease α-sarcin to determine the binding sites for *Escherichia coli* ribosomal proteins L5, L18, and L25 on 5 S rRNA. *Proc. Natl. Acad. Sci. USA* **81**, 322-326.

Ikura, M., Kay, L. E. & Bax, A. (1990). A novel approach for sequential assignment of ^{1}H, ^{13}C, and ^{15}N spectra of larger proteins: heteronuclear triple resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry* **29**, 4659-4667.

Innis, M. A., Gelfand, D. H., Sninsky, J. J. & White, T. J., Eds. (1990). PCR Protocols: A Guide to Methods and Applications. New York: Academic Press, Inc.

Jack, A., Lander, J. E. & Klug, A. (1976). Crystallographic refinement of yeast phenylalanine transfer RNA at 2.5 Å resolution. *J. Mol. Biol.* **108**, 619-649.

Jacob, W. F., Santer, M. & Dahlberg, A. E. (1987). A single base change in the Shine-Dalgarno region of 16S rRNA of *Escherichia coli* affects translation of many proteins. *Proc. Natl. Acad. Sci. USA* **84**, 4757-4761.

Jaeger, L., Michel, F. & Westhof, E. (1994). Involvement of a GNRA tetraloop in long-range RNA tertiary interactions. *J. Mol. Biol.* **236**, 1271-.

Jaishree, T. N., Ramakrishnan, V. & White, S. W. (1996). Solution structure of prokaryotic ribosomal protein S17 by high resolution NMR spectroscopy. *Biochemistry* **35**, 2845-2853.

Jessen, T.-H., Oubridge, C., Teo, C. H., Pritchard, C. & Nagai, K. (1991). Identification of molecular contacts between the U1A small nuclear ribonucleoprotein and U1 RNA. *EMBO J.* **10**, 3447-3456.

Joshua-Tor, L., Rabinovich, D., Hope, H., Frolow, F., Appella, E. & L., S. J. (1988). The three-dimensional structure of a DNA duplex containing looped out bases. *Nature* **334**, 82-84.

Jucker, F. M. & Pardi, A. (1995). Solution structure of the CUUG hairpin loop: A novel RNA tetraloop motif. *Biochemistry* **34**, 14416-14427.

Jutta, R.-A., Junke, N., Brimacombe, R., Lavrik, I., Dokudovskaya, S., Dontsova, O. & Bogdanov, A. (1994). Contacts between 16S ribosomal RNA and mRNA, within the spacer region separating the AUG initiator codon and the Shine-Dalgarno sequence; a site-directed cross-linking study. *Nucleic Acids Res.* **22**, 3018-3025.

Juzumiene, D. I., Shapkina, T. G. & Wollenzien, P. (1995). Distribution of cross-links between mRNA analogues and 16S rRNA in *Escherichia coli* 70S ribosomes made under equilibrium conditions and their response to tRNA binding. *J. Biol. Chem.* **270**, 12794-12800.

Kamp, R. M., Bosserhoff, A., Kamp, D. & Wittman-Liebold, B. (1984). Application of High-Performance Liquid Chromatographic techniques to the separation of ribosomal proteins of different organisms. *J. Chrom.* **317**, 181-192.

Karn, J. (1991). Control of human immunodeficiency virus replication by the tat, rev, nef and protease genes. *Curr. Opin. Immun.* **3**, 526-536.

Katahira, M., Kanagawa, M., Sato, H., Uesugi, S., Fujii, S., Kohno, T. & Maeda, T. (1994). Formation of sheared G:A base pairs in an RNA duplex modelled after ribozymes, as revealed by NMR. *Nucleic Acids Res.* **22**, 2752-2759.

Kearny, K. R. & Nomura, M. (1987). Secondary structure of the autoregulatory mRNA binding site of ribosomal protein L1. *Mol. Gen. Genet.* **210**, 60-68.

Khechinashvili, N. N., Koteliansky, V. E., Gogia, Z. V., Littlechild, J. & Dijk, J. (1978). A heat denaturation study of ribosomal proteins from *Escherichia coli* by scanning microcalorimetry. *Fed. Eur. Biochem. Soc. Lett.* **95**, 270-272.

Kjems, J., Calnan, B., Frankel, A. D. & Sharp, P. A. (1992). Specific binding of a basic peptide form HIV-1 Rev. *EMBO J.* **11**, 1119-1129.

Knapp, G. (1989). Enzymatic approaches to probing RNA secondary and tertiary structure. *Methods Enzymol.* **180**, 192-212.

Konisky, J. & Nomura, M. (1967). Interaction of colicins with bacterial cells. II. Specific alteration of *E. coli* ribosomes induced by colicin E3 *in vivo*. *J. Mol. Biol.* **26**, 181-195.

Kooi, E. A., Rutgers, C. A., Mulder, A., Riet, J. V. t., Venema, J. & Raue, H. A. (1993). The phylogenetically conserved doublet tertiary interaction in domain III of the large subunit rRNA is crucial for ribosomal protein binding. *Proc. Natl. Acad. Sci. USA* **90**, 213-216.

Krauss, G., Romer, R., Riesner, D. & Maass, G. (1973). Thermodynamics and kinetics of the interactin of phenylalanine-specific tRNA from yeast with its cognate synthetase as studied by the fluorescence of the Y-base. *FEBS Lett.* **30**, 6-10.

Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E. & Cech, T. R. (1982). Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell* **31**, 147-157.

Lambowitz, A. M. & Perlman, P. S. (1990). Involvement of aminoacyl-tRNA synthetases and other proteins in group I and II intron splicing. *Trends Biochem. Sci.* **15**, 440-444.

Lawley, P. D. & Brookes, P. (1963). Further studies on the alkylation of nucleic acids and their constituent nucleotides. *Biochem. J.* **89**, 117-138.

Lee, D. K., Horikoshi, M. & Roeder, R. G. (1991). Interaction of TFIID in the minor groove of the TATA element. *Cell* **67**, 1241-1250.

Legrain, P. & Rosbash, M. (1989). Some cis- and trans-acting mutants for splicing target pre-mRNA to the cytoplasm. *Cell* **57**, 573-583.

LeMaster, D. M. (1989). Deuteration in protein proton magnetic resonance. *Methods Enzymol.* **177**, 23-43.

LeMaster, D. M. (1990). Deuterium labeling in NMR structural analysis of larger proteins. *Q. Rev. Biophys.* **23**, 133-174.

Leonard, N. J., McDonald, J. J., Henderson, R. E. L. & Reichmann, D. E. (1971). Reaction of diethyl pyrocarbonate with nucleic acid components. Guanine, cytosine, and uracil. *Biochemistry* **10**, 3335-3342.

Leontis, N. B., Kwok, W. & Newman, J. S. (1991). Stability and structure of three-way DNA junctions containing unpaired nucleotides. *Nucleic Acids Res.* **19**, 759-766.

Lilley, D. M. J. (1994). Homologous recombination. A ring for a warhead. *Curr. Biol.* **4**, 1152-1154.

Lim, C. K. & Peters, J. J. (1989). Isocratic reverse-phase high performance liquid chromatography of ribonucleotides, deoxyribonucleotides, cyclic nucleotides and cyclic deoxyribonucleotides. *J. Chrom.* **461**, 259-266.

Lin, S.-y. & Riggs, A. D. (1972). *lac* repressor binding to non-operator DNA: detailed studies and a comparison of equilibrium and rate competition methods. *J. Mol. Biol.* **72**, 671-690.

Lindahl, M., Svensson, L. A., Liljas, A., Sedelnikova, S. E., Eliseikina, I. A., Fomenkova, N. P., Nevskaya, N., Nikonov, S. V., Garber, M. B., Muranova, T. A., Rykonova, A. I. & Amons, R. (1994). Crystal structure of the ribosomal protein S6 from *Thermus thermophilus*. *EMBO J.* **13**, 1249-1254.

Long, A. R. & Anthony, C. (1990). Modifier protein for methanol dehydrogenase of methylotrophs. *Methods Enzymol.* **188**, 216-222.

Long, D. M. & Uhlenbeck, O. C. (1993). Self-cleaving catalytic RNA. *FASEB J.* **7**, 25-30.

Long, K. S. & Crothers, D. M. (1995). Interaction of Human Immunodeficiency Virus type 1 tat-derived peptides with TAR RNA. *Biochemistry* **34**, 8885-8895.

Lutcke, H. (1995). Signal recognition particle (SRP), a ubiquitous initiator of protein translocation. *Eur. J. Biochem.* **228**, 531-550.

Madhani, H. D. & Guthrie, C. (1994). Dynamic RNA-RNA interactions in the spliceosome. In *Annual Review of Genetics*, Vol. 28, pp. 1-26. Annual Reviews, Inc.

Malhotra, A. & Harvey, S. C. (1994). A quantitative model of the *Escherichia coli* 16S RNA in the 30S ribosomal subunit. *J. Mol. Biol.* **240**, 308-340.

Mandiyan, V., Tumminia, S., Wall, J. S., Hainfeld, J. F. & Boublik, M. (1989). Protein-induced conformational changes in 16 S ribosomal RNA during the initial assembly steps of the *Escherichia coli* 30 S ribosomal subunit. *J. Mol. Biol.* **210**, 323-336.

Marino, J. P., Schwalbe, H., Anklin, C., Bermel, W., Crothers, D. M. & Griesinger, C. (1994). A three-dimensional triple-resonance $^{1}H,^{13}C,^{31}P$ experiment: sequential through-bond correlation of ribose protons and intervening phosphorous along the RNA oligonucleotide backbone. *J. Am. Chem. Soc.* **116**, 6472-6473.

Maxam, A. M. & Gilbert, W. (1977). A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA* **74**, 560-564.

Mazzeo, J. R. & Krull, I. S. (1989). Immobilized boronates for the isolation and separation of bioanalytes. *BioChromatography* **4**, 124-130.

McClain, W. H. & Foss, K. (1988). Changing the identity of a tRNA by introducing a G-U wobble pair near the 3' acceptor end. *Science* **240**, 793-796.

Meyer, B. & Malim, M. H. (1994). The HIV-1 Rev *trans*-activator shuttles between the nucleus and the cytoplasm. *Genes Dev.* **8**, 1538-1547.

Michnicka, M. J., Harper, J. W. & King, G. C. (1993). Selective isotopic enrichment of synthetic RNA : Application to the HIV-1 TAR element. *Biochemistry* **32**, 395-400.

Miller, J. H. (1972). *Experiments in Molecular Genetics*. Experiments in Molecular Genetics, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Miller, M., Harrison, R. W., Wlodawer, A., Appella, E. & Sussman, J. L. (1988). Crystal structure of 15-mer DNA duplex containing unpaired bases. *Nature* **334**, 85-86.

Milligan, J. F., Groebe, D. R., Witherell, G. W. & Uhlenbeck, O. C. (1987). Oligoribonucleotide synthesis using T7 RNA-polymerase and synthetic DNA templates. *Nucleic Acids Res.* **15**, 8783-8798.

Mizushima, S. & Nomura, M. (1970). Assembly mapping of 30S ribosomal proteins from *E. coli*. *Nature* **226**, 1214-1218.

Moazed, D. & Noller, H. F. (1986). Transfer RNA shields nucleotides in 16S ribosomal RNA from attack by chemical probes. *Cell* **47**, 985-994.

Moazed, D. & Noller, H. F. (1987). Interaction of antibiotics with functional sites in 16S ribosomal RNA. *Nature* **327**, 389-394.

Moazed, D. & Noller, H. F. (1990). Binding of tRNA to the ribosomal A and P site protects two distinct sets of nucleotides in 16S rRNA. *J. Mol. Biol.* **211**, 135-145.

Moazed, D., Stern, S. & Noller, H. F. (1986). Rapid chemical probing of conformation in 16 S ribosomal RNA and 30 S ribosomal subunits using primer extension. *J. Mol. Biol.* **187**, 399-416.

Molinaro, M. & Tinoco, I., Jr. (1995). Use of ultra stable UNCG tetraloop hairpins to fold RNA structures: thermodynamic and spectroscopic applications. *Nucleic Acids Res.* **23**, 3056-3063.

Moore, M. J., Query, C. C. & Sharp, P. A. (1993). Splicing of precursors to mRNA by the spliceosome. In *The RNA world* (Gesteland, R. & Atkins, J. F., eds.), pp. 137-156. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Moore, P. B. (1979). The preparation of deuterated ribosomal materials for neutron scattering. *Methods Enzymol.* **59**, 639-655.

Moras, D., Comarmond, M. B., Fischer, J., Weiss, R., Thierry, J. C., Ebel, J. P. & Giege, R. (1980). Crystal structure of yeast $tRNA^{Asp}$. *Nature* **288**, 669-674.

Morgan, J. & Brimacombe, R. (1972). A series of specific ribonucleoprotein fragments from the 30S subparticle of *Escherichia coli* ribosomes. *Eur. J. Biochem.* **29**, 542-552.

Morinaga, T., Funatsu, G., Funatsu, M. & Wittman, H. G. (1976). Primary structure of the 16S rRNA binding protein S15 from *Escherichia coli* ribosomes. *FEBS Lett.* **64**, 307-309.

Morrison, C. A., Bradbury, E. M. & Garrett, R. A. (1977). A comparison of the structures of several acid-urea extracted ribosomal proteins from *Escherichia coli* using proton NMR. *Fed. Eur. Biochem. Soc. Lett.* **81**, 435-439.

Mougel, M., Allmang, C., Eyermann, F., Cachia, C., Ehresmann, B. & Ehresmann, C. (1993). Minimal 16S rRNA binding site and role of conserved nucleotides in *Escherichia coli* ribosomal protein S8 recognition. *Eur. J. Biochem.* **215**, 787-792.

Mougel, M., Ehresmann, B. & Ehresmann, C. (1986). Binding of *Escherichia coli* ribosomal protein S8 to 16S rRNA: kinetic and thermodynamic characterization. *Biochemistry* **1986**, 2756-2765.

Mougel, M., Eyermann, F., Westhof, E., Romby, P., Expert-Bezancon, A., Ebel, J.-P., Ehresmann, B. & Ehresmann, C. (1987). Binding of *Escherichia coli* ribosomal protein S8 to 16S rRNA. *J. Mol. Biol.* **198**, 91-107.

Mougel, M., Philippe, C., Ebel, J., Ehresmann, B. & Ehresmann, C. (1988). The *E. coli* 16S rRNA binding site of ribosomal protein S15: higher-order structure in the absence and in the presence of the protein. *Nucleic Acids Res.* **16**, 2825-2839.

Müller, R., Garrett, R. A. & Noller, H. F. (1979). The structure of the RNA binding site of ribosomal proteins S8 and S15. *J. Biol. Chem.* **254**, 3873-3878.

Murchie, A. I. H., Clegg, R. M., von Kitzing, E., Duckett, D. R., Diekmann, S. & Lilley, D. M. J. (1989). Fluorescence energy transfer shows that the four-way DNA junction is right-handed cross of antiparallel molecules. *Nature* **341**, 733-736.

Murgola, E. J., Hijazi, K. A., Goringer, H. U. & Dahlberg, A. E. (1988). Mutant 16S ribosomal RNA: a codon-specific translational suppressor. *Proc. Natl. Acad. Sci. USA* **85**, 4162-4165.

Musier-Forsyth, K. & Schimmel, P. (1992). Functional contacts of a transfer RNA synthetase with 2'-hydroxyl groups in the RNA minor groove. *Nature* **357**, 513-515.

Musier-Forsyth, K., Usman, N., Scaringe, S., Doudna, J., Green, R. & Schimmel, P. (1991). Specificity for aminoacylation of a RNA helix: An unpaired, exocyclic amino group in the minor groove. *Science* **253**, 784-786.

Neidhardt, F. C. (1987). Chemical compostion of *Escherichia coli*. In *Escherichia coli and Salmonella typhimurium. Cellular and Molecular Biology.* (Neidhardt, F. C., ed.), Vol. 1, pp. 2-6. 2 vols. American Society for Microbiology, Washington, D.C.

Nietlispach, D., Clowes, R. T., Broadhurst, R. W., Ito, Y., Keeler, J., Kelly, M., Ashurst, J., Oschkinat, H., Domaille, P. J. & Laue, E. D. (1996). An approach to the structure determination of larger proteins using triple resonance NMR experiments in conjunction with random fractional deuteration. *J. Am. Chem. Soc.* **118**, 407-415.

Nikonowicz, E. P. & Pardi, A. (1992a). Application of four-dimensional heteronuclear NMR to the structure determination of a uniformly ^{13}C labeled RNA. *J. Am. Chem. Soc.* **114**, 1082-1083.

Nikonowicz, E. P. & Pardi, A. (1992b). Three-dimensional heteronuclear NMR studies of RNA. *Nature* **355**, 184-186.

Nikonowicz, E. P. & Pardi, A. (1993). An efficient procedure for assignment of the proton, carbon and nitrogen resonances in carbon-13/nitrogen-15 labeled nucleic acids. *J. Mol. Biol.* **232**, 1141-1156.

Nikonowicz, E. P., Sirr, A., Legault, P., Jucker, F. M., Baer, L. M. & Pardi, A. (1992). Preparation of carbon-13- and nitrogen-15-labeled RNAs for heteronuclear multidimensional NMR studies. *Nucleic Acids Res.* **20**, 4507-4513.

Nissen, P., Kjeldgaard, M., Thirup, S., Polekhina, G., Reshetnikova, L., Clark, B. F. C. & Nyborg, J. (1995). Crystal structure of the ternary complex of Phe-tRNA^{Phe}, EF-Tu, and a GTP analog. *Science* **270**, 1464-1472.

Noller, H. F. (1993). On the origin of the ribosome: Coevolution of subdomains of tRNA and rRNA. In *The RNA world* (Gesteland, R. & Atkins, J. F., eds.), pp. 137-156. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Noller, H. F., Moazed, D., Stern, S., Powers, T., Allen, P. N., Robertson, J. M., Weiser, B. & Triman, K. (1989). Structure of rRNA and its functional interactions in translation. In *The Ribosome: Structure, function, and evolution* (Hill, W. E., Dahlberg, A., Garrett, R. A., Moore, P. B., Schlessinger, D. & Warner, J. R., eds.), pp. 73-92. American Society for Microbiology, Washington D.C.

Noller, H. F. & Nomura, M. (1987). Ribosomes. In *Escherichia coli and Salmonella Typhimurium. Cellular and molecular biology* (Neidhardt, F. C., ed.), Vol. Vol. I, pp. 104-125. American Society for Microbiology, Washington, D. C.

Noller, H. F. & Woese, C. R. (1981). Secondary Structure of 16S Ribosomal RNA. *Science* **212**, 403-411.

Nomura, M. (1963). Mode of action of colicines. *Cold Spring Harbor Symp. Quant. Biol.* **28**, 315-324.

Nomura, M., Traub, P. & Bechmann, H. (1968). Hybrid 30S ribosomal particles reconstituted from components of different bacterial origins. *Nature* **219**, 793-799.

Nomura, M., Traub, P., Guthrie, C. & Nashimoto, H. (1969). The assembly of ribosomes. *J. Cell Physiol.* **74**, 241-252.

Nomura, M., Yates, J. L., Dean, D. & Post, L. E. (1980). Feedback regulation of ribosomal protein gene expression in *Escherichia coli*: structural homology of ribosomal RNA and ribosomal protein mRNA. *Proc. Natl. Acad. Sci. USA* **77**, 7084-7088.

Nowotny, V. & Nierhaus, K. H. (1988). Assembly of the 30S subunit from *Escherichia coli* ribosomes occurs via two assembly domains which are initiated by S4 and S7. *Biochemistry* **27**, 7051-7055.

O'Conner, M., Goringer, H. U. & Dahlberg, A. E. (1992). A ribosomal ambiguity mutation in the 530 loop of *E. coli* 16S rRNA. *Nucleic Acids Res.* **87**, 4221-4227.

Oakley, M. G. & Dervan, P. B. (1990). Structural motif of the GCN4 DNA binding domain characterized by affinity cleaving. *Science* **248**, 847-850.

Oliphant, A. R., Brandl, C. J. & Struhl, K. (1989). Defining the sequence specificity of DNA-binding proteins by selecting binding sites from random-sequence oligonucleotides: anaylsis of yeast GCN4 protein. *Mol. Cell. Biol.* **9**, 2944-2949.

Oubridge, C., Ito, N., Evans, P. R., Teo, C.-H. & Nagai, K. (1994). Crystal structure at1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **372**, 432-438.

Pan, T., Long, D. M. & Uhlenbeck, O. C. (1993). Divalent metal ions in RNA folding and catalysis. In *The RNA world* (Gesteland, R. & Atkins, J. F., eds.), pp. 271-302. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Parker, S. A., Batey, R. T. & Williamson, J. R. unpublished observations.

Patel, D. J., Kozlowski, S. A., Marky, L. A., Rice, J. A., Broka, C., Itakura, K. & Breslauer, K. J. (1982). Extra adenosine stacks into the self complimentary d(CGCAGAATTCGCG) duplex in solution. *Biochemistry* **21**, 445-451.

Peattie, D. A. (1979). Direct chemical method for sequencing RNA. *Proc. Natl. Acad. Sci. USA* **76**, 1760-1764.

Peattie, D. A. & Gilbert, W. (1980). Chemical probes for higher-order structure in RNA. *Proc. Natl. Acad. Sci. USA* **77**, 4679-4682.

Philippe, C., Benard, L., Eyermann, F., Cachia, C., Kirillov, S. V., Portier, C., Ehresmann, B. & Ehresmann, C. (1994). Structural elements of *rpsO* mRNA

involved in the modulation of translational initiation and regulation of *E. coli* ribosomal protein S15. *Nucleic Acids Res.* **22**, 2538-2546.

Philippe, C., Portier, C., Mougel, M., Grunberg-Manago, M., Ebel, J. P., Ehresmann, B. & Ehresmann, C. (1990). Target site of *Escherichia coli* ribosomal protein S15 on its messenger RNA. *J. Mol. Biol.* **211**, 415-426.

Phillipe, C., Benard, L., Portier, C., Westhof, E., Ehresmann, B. & Ehresmann, C. (1995). Molecular dissection of the pseudoknot governing the translational regulation of *Escherichia coli* ribosomal protein S15. *Nucleic Acids Res.* **23**, 18-28.

Phillipe, C., Eyermann, F., Benard, L., Portier, C., Ehresmann, B. & Ehresmann, C. (1993). Ribosomal protein S15 from *Escherichia coli* modulates its own translation by trapping the ribosome on the mRNA initiation loading site. *Proc. Natl. Acad. Sci. USA* **90**, 4394-4398.

Pingoud, A., Boehme, D., Riesner, D., Kownatzki, R. & Maass, G. (1975). Anti-Cooperative binding of two $tRNA^{Tyr}$ Molecules to tyrosyl-tRNA synthetase from *Escherichia coli*. *Eur. J. Biochem.* **56**, 617-622.

Pingoud, A., Riesner, D., Boehme, D. & Maass, G. (1973). Kinetic studies on the interaction of seryl-tRNA synthetase with $tRNA^{Ser}$ and Ser-$tRNA^{Ser}$ from yeast. *FEBS Lett.* **30**, 1-5.

Pley, H., Flaherty, K. & McKay, D. (1994a). Model for an RNA tertiary interaction form the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix. *Nature* **372**, 111-113.

Pley, H. W., Flaherty, K. M. & McKay, D. B. (1994b). Three-dimensional structure of a hammerhead ribozyme. *Nature* **372**, 68-74.

Polson, A. G., Crain, P. F., Pomerantz, S. C., McCloskey, J. A. & Bass, B. L. (1991). The mechanism of adenosine to inosine conversion by the double-stranded RNA unwinding/modifying activity: A high-performance liquid chromatography-mass spectrometry analysis. *Biochemistry* **30**, 11507-11514.

Poritz, M. A., Bernstein, H. D., Strub, K., Zopf, D., Wilhelm, H. & Walter, P. (1990). An *E. coli* ribonucleoprotein containing 4.5S RNA resembles mammalian signal recognition particle. *Science* **250**, 1111-1117.

Portier, C., Dondon, L. & Grunberg-Manago, M. (1990a). Translational autocontrol of the *Escherichia coli* ribosomal protein S15. *J. Mol. Biol.* **211**, 407-417.

Portier, C., Dondon, L. & Grunberg-Manango, M. (1990b). Translational autocontrol of the *Escherichia coli* ribosomal protein S15. *J. Mol. Biol.* **211**, 407-414.

Portier, C., Philippe, E., Dondon, L., Grunberg-Manago, M., Ebel, J. P., Ehresmann, B. & Ehresmann, C. (1990c). Translational control of ribosomal protein S15. *Biochimica et Biophysica Acta* **1050**, 328-336.

Powers, T., Changchien, L.-M., Craven, G. R. & Noller, H. F. (1988a). Probing the assembly of the 3' major domain of 16S ribosomal RNA: Quarternary interactions involving ribosomal proteins S7, S9 and S19. *J. Mol. Biol.* **200**.

Powers, T. & Noller, H. F. (1990). Dominant lethal mutations in a conserved loop in 16S rRNA. *Proc. Natl. Acad. Sci. USA* **87**, 1042-1046.

Powers, T. & Noller, H. F. (1991). A functional pseudoknot in 16S ribosomal RNA. *EMBO J.* **10**, 2203-2214.

Powers, T. & Noller, H. F. (1995). Hydroxyl radical footprinting of ribosomal proteins on 16S rRNA. *RNA* **1**, 194-209.

Powers, T., Stern, S., Changchien, L.-M. & Noller, H. F. (1988b). Probing the assembly of the 3' major domain of the 16S ribosomal RNA: Interactions involving ribosomal proteins S2, S3, S10, S13, and S14. *J. Mol. Biol.* **201**, 697-716.

Predki, P. F., Nayak, L. M., Gottlieb, M. B. C. & Regan, L. (1995). Dissecting RNA-protein interactions: RNA-RNA recognition by rop. *Cell* **80**, 41-50.

Price, S. R., Ito, N., Oubridge, C., Avis, J. M. & Nagai, K. (1995). Crystallization of RNA-protein complexes I. Methods for the large-scale preparation of RNA suitable for crystallographic studies. *J. Mol. Biol.* **249**, 398-408.

Prince, J. B., Taylor, B. H., Thurlow, D. L., Ofengand, J. & Zimmerman, R. A. (1982). Covalent crosslinking of $tRNA^{Val}$ to 16S RNA at the ribosomal P site: Identification of crosslinked sites. *Proc. Natl. Acad. Sci. USA* **79**, 5450-5454.

Pritchard, C. E., Grasby, J. A., Hamy, F., Zacharek, A. M., Singh, M., Karn, J. & Gait, M. (1994). Methylphosphonate mapping of phosphate contacts critical for RNA recognition by the human immunodeficiency virus tat and rev proteins. *Nucleic Acids Res.* **22**, 2592-2600.

Puglisi, J. D. (1989). RNA folding: Structure and conformational equilibria of RNA pseudoknots. Doctoral Thesis, University of California at Berkeley.

Puglisi, J. D., Chen, L., Blanchard, S. & Frankel, A. D. (1995). Solution structure of a Bovine Immunodeficiency Virus tat-TAR peptide-RNA complex. *Science* **270**, 1200-1203.

Puglisi, J. D., Tan, R., Calnan, B. J., Frankel, A. D. & Williamson, J. R. (1992). Conformation of the TAR RNA-arginine complex by NMR spectroscopy. *Science* **257**, 76-80.

Puglisi, J. D. & Tinoco, I., Jr. (1989). Absorbance melting curves of RNA. In *Methods in Enzymology*, Vol. 180, pp. 304-325.

Puglisi, J. D., Wyatt, J. R. & Tinoco, I., Jr. (1990). Conformation of an RNA pseudoknot. *J. Mol. Biol.* **214**, 437-453.

Rabi, J. A. & Fox, J. J. (1973). Nucleosides. LXXXIX. Facile base-catalyzed hydrogen isotope labeling at position 6 of pyrimidine nucleosides. *J. Am. Chem. Soc.* **95**, 1628-1632.

Ramakrishnan, V. & White, S. W. (1992). The structure of ribosomal protein S5 reveals sites of interaction with 16S rRNA. *Nature* **358**, 768-771.

Recht, M. I., Fourmy, D., Blanchard, S. C., Dahlquist, K. D. & Puglisi, J. D. (1996). RNA sequence determinants for aminoglycoside binding to an A-site rRNA model oligonucleotide. *J. Mol. Biol.* **262**, 421-436.

Reddy, K. J. & Gilman, M. (1989). Preparation of Bacterial RNA. In *Current Protocols in Molecular Biology* (Ausubel, F. M., Brent, R. E., Kingston, D. D., Moore, J. G., Seidman, J. A., Smith, J. A. & Struhl, K., eds.), pp. 4.4.1-7. John Wiley & Sons, New York.

Revzin, A. (1989). Gel electrophoresis assays for DNA-protein interactions. *Biotechniques* **7**, 346-355.

Romaniuk, P. J. (1989). The role of highly conserved single-stranded nucleotides of *Xenopus* 5 S RNA in the binding of transcription factor IIIA. *Biochemistry* **28**, 1388-1395.

Rould, M. A., Perona, J. J., Soll, D. & Steitz, T. A. (1989). Structure of *E. Coli* glutaminyl-tRNA synthetase complexed with $tRNA^{Gln}$ and ATP at 2.8 Å resolution: implications for tRNA discrimination. *Science* **246**, 1135-1142.

Roy, S., Sklenar, V., Appella, E. & Chen, J. S. (1987). Conformational perturbation due to an extra adenosine in a self-complementary oligodeoxynucleotide. *Biopolymers* **26**, 2041-2052.

Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J. C. & Moras, D. (1991). Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with $tRNA^{Asp}$. *Science* **252**, 1682-1689.

Ryan, P. C. & Draper, D. E. (1989). Thermodynamics of protein-RNA recognition in a highly conserved region of the large-subunit ribosomal RNA. *Biochemistry* **28**, 9949-9956.

Ryan, P. C., Lu, M. & Draper, D. E. (1991). Recognition of the highly conserved GTPase center of 23 S ribosomal RNA by ribosomal protein L11 and the antibiotic thiostrepton. *J. Mol. Biol.* **221**, 1257-1268.

Saldanha, R. J., Patel, S. S., Surendran, R., Lee, J. C. & Lambowitz, A. M. (1995). Involvement of *Neurospora* mitochondrial tyrosyl-tRNA synthetase in RNA splicing. A new method for purifying the protein and characterization of physical and enzymatic properties pertinent to splicint. *Biochemistry* **34**, 1275-1287.

Samaha, R. R., O'Brien, B., O'Brien, T. W. & Noller, H. F. (1994). Independent *in vitro* assembly of a ribonucleoprotein particle containing the 3' domain of 16S rRNA. *Proc. Natl. Acad. Sci. USA* **91**, 7884-7888.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual.* Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

SantaLucia, J. J. & Turner, D. H. (1993). Structure of (rGGCGAGCC)$_2$ in solution from NMR and restrained molecular dynamics. *Biochemistry* **32**, 12612-12623.

Santer, M., Bennet-Guerrero, E., Byahatti, S., Czarnecki, S., O'Connell, D., Meyer, M., Khoury, J., Cheng, X., Schwartz, I. & McLaughlin, J. (1990). Base changes at position 792 of *Escherichia coli* 16S rRNA affect assembly of 70S ribosomes. *Proc. Natl. Acad. Sci. USA* **87**, 3700-3704.

Santoro, J. & King, G. C. (1992). A constant-time 2D Overbodenhausen experiment for inverse correlation of isotopically enriched species. *J. Mag. Reson.* **97**, 202-207.

Sattler, M. & Fesik, S. W. (1996). Use of deuterium labeling in NMR: overcoming a sizeable problem. *Structure* **4**, 1245-1249.

Schatz, D., Leberman, R. & Eckstein, F. (1991). Interaction of *Escherichia coli* tRNA^{Ser} with its cognate aminoacyl-tRNA synthetase as determined by footprinting with phophorothioate-containing tRNA transcripts. *Proc. Natl. Acad. Sci. USA* **88**, 6132-6136.

Schevitz, R. W., Podjarny, A. D., Krishnamachari, N., Hughes, J. J., Sigler, P. B. & Sussman, J. L. (1979). Crystal structure of a eukaryotic initiator tRNA. *Nature* **248**, 188-190.

Schott, H., Rudloff, E., Schmidt, P., Roychoudhury, R. & Kossel, H. (1973). A dihydroxyboryl-substituted methacrylic polymer for the column chromatographic separation of mononucleotides, oligonucleotides, and transfer ribonucleic acid. *Biochemistry* **12**, 932-938.

Schwarzbauer, J. & Craven, G. R. (1981). Apparent association constants for *E. coli* ribosomal proteins S4, S7, S8, S15, S17 and S20 binding to 16S RNA. *Nucleic Acids Res.* **9**, 2223-2237.

Scott, W. G., Finch, J. T., Grenfell, F., Fogg, J., Smith, T., Gait, M. J. & Klug, A. (1995a). Rapid crystallization of chemically synthesized hammerhead RNAs using a double screening procedure. *J. Mol. Biol.* **250**, 327-332.

Scott, W. G., Finch, J. T. & Klug, A. (1995b). The crystal structure of an all-RNA hammerhead ribozyme: A proposed mechanism for RNA catalytic cleavage. *Cell* **81**, 991-1002.

Scott, W. G. & Klug, A. (1996). Ribozymes: structure and mechanism in RNA catalysis. *Trends Biochem. Sci.* **21**, 220-223.

Seeman, N. C. & Kallenbach, N. R. (1994). DNA branched junctions. In *Annual Review of Biophysics and Biomolecular Structure*, Vol. 23, pp. 53-86.

Senear, D. F. & Brenowitz, M. (1991). Determination of binding constants for cooperative site-specific protein-DNA interactions using the gel mobility-shift assay. *J. Biol. Chem.* **266**, 13661-13671.

Serganov, A. A., Masquida, B., Westhof, E., Cachia, C., Portier, C., Garber, M., Ehresmann, B. & Ehresmann, C. (1996). The 16S rRNA binding site of *Thermus*

thermophilus ribosomal protein S15: Comparison with *Escherichia coli* S15, minimum site and structure. *RNA* **2**, 1124-1138.

Shabarova, Z. & Bogdanov, A. (1994). *Advanced Organic Chemistry of Nucleic Acids.* First edit, VCH Verlagsgesellschaft mbH, New York.

Shen, L. X. & Tinoco, I., Jr. (1995). The structure of an RNA pseudoknot that causes efficient frameshifting in mouse mammary tumor virus. *J. Mol. Biol.* **247**, 963-978.

Shen, Z. & Hagerman, P. J. (1994). Conformation of the central, three-helix junction of the 5 S ribosomal RNA of *Sulfolobus acidocaldarius. J. Mol. Biol.* **241**, 415-430.

Shine, J. & Dalgarno, L. (1974). The 3'-terminal sequence of *E. coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. USA* **71**, 1342-1346.

Simon, E. S., Grabowski, S. & Whitesides, G. M. (1990). Convenient syntheses ofcytidine 5'-triphosphate, guanosine 5'-triphosphate, and uridine 5'-triphosphate and their use in the preparation of UDP-glucose, UDP-glucoronic acid, and GDP-mannose. *J. Org. Chem.* **55**, 1834-1841.

Sklenar, V., Peterson, R. D., Rejante, M. R. & Feigon, J. (1994). Correlation of nucleotide base and sugar protons in a ^{15}N-labeled HIV-1 RNA oligonucleotide by ^{1}H-^{15}N HSQC experiments. *J. Biol. NMR* **4**, 117-122.

Sklenar, V., Peterson, R. D., Rejante, M. R., Wang, E. & Feigon, J. (1993). Two-Dimensional Triple-Resonance HCNCH Experiment for Direct Correlation of Ribose H1' and Base H8, H6 Protons in 13C,15N-Labeled RNA Oligonucleotides. *J. Am. Chem. Soc.* **115**, 12181-12182.

Spierer, P., Bodgdanov, A. A. & Zimmerman, R. A. (1978). Parameters for the interaction of ribosomal proteins L5, L18, and L25 with 5S RNA from *Escherichia coli. Biochemistry* **17**, 5394-5398.

Stark, M. J. R., Gregory, R. J., Gourse, R. L., Thurlow, D. L., Zwieb, C., Zimmerman, R. A. & Dahlberg, A. E. (1984). Effects of site-directed mutations in the central domain of 16 S ribosomal RNA upon ribosomal protein binding, RNA processing and 30 S aubunit assembly. *J. Mol. Biol.* **178**, 303-322.

Steitz, J. A. (1992). Splicing takes a Holliday. *Science* **257**, 888-889.

Steitz, J. A. & Jakes, K. (1975). How ribosomes select intiator regions in mRNA: base pair formation between the 3' terminus of 16S rRNA and the mRNA during initiation of protein synthesis in *E. coli. Proc. Natl. Acad. Sci. USA* **72**, 4734-4738.

Stern, S., Changchien, L.-M., Craven, G. R. & Noller, H. F. (1988a). Interaction of proteins S16, S17 and S20 with 16S ribosomal RNA. *J. Mol. Biol.* **200**, 291-299.

Stern, S., Powers, T., Changchien, L.-M. & Noller, H. F. (1988b). Interaction of ribosomal proteins S5, S6, S11, S12, S18 and S21 with 16S rRNA. *J. Mol. Biol.* **201**, 683-695.

Stern, S., Weiser, B. & Noller, H. F. (1988c). Model for the three-dimensional folding of 16 S ribosomal RNA. *J. Mol. Biol.* **204**, 447-481.

Stern, S., Wilson, R. C. & Noller, H. F. (1986). Localization of the binding site for protein S4 on 16S ribosomal RNA by chemical and enzymatic probing and primer extension. *J. Mol. Biol.* **192**, 101-110.

Stirling, D. I. (1987). Eur. Pat. Appl. #231585.

Svensson, P., Changchien, L., Craven, G. R. & Noller, H. F. (1988). Interaction of ribosomal proteins, S6, S8, S15, and S18 with the central domain of 16S ribosomal RNA. *J. Mol. Biol.* **200**, 301-308.

Sylvers, L. A., Kopylov, A., Wower, J., Hixson, S. S. & Zimmerman, R. A. (1992). Photochemical cross-linking of the anticodon loop of yeast $tRNA^{Phe}$ to 30S-subunit protein S7 at the ribosomal A and P sites. *Biochimie* **74**, 381-389.

Szewczak, A. A. & Moore, P. B. (1995). The sarcin/ricin loop, a modular RNA. *J. Mol. Biol.* **247**, 81-98.

Szewczak, A. A., Moore, P. B., Chan, Y.-L. & Wool, I. G. (1993). The conformation of the sarcin/ricin loop from 28S ribosomal RNA. *Proc. Natl. Acad. Sci. USA* **90**, 9581-9585.

Tao, F., Batey, R. T. & Williamson, J. R. unpublished observations.

Tao, F., Parker, S. A., Batey, R. T. & Williamson, J. R. unpublished observations.

Tao, J. & Frankel, A. D. (1992). Specific binding of arginine to TAR RNA. *Proc. Natl. Acad. Sci. USA* **89**, 2723-2726.

Tapprich, W. E., Goss, D. J. & Dahlberg, A. E. (1989). Mutation at position 791 in *Escherichia coli* 16S riboomal RNA affects processes involved in the initiation of protein synthesis. *Proc. Natl. Acad. Sci. USA* **86**, 4927-4931.

Temperli, A., Turler, H., Rust, P., Danon, A. & Chargaff, E. (1964). Studies on the nucleotide arrangement in deoxyribonucleic acids. IX. Selective degradation of pyrimidine deoxyribonucleotides. *Biochimica et Biophysica Acta* **91**, 462-476.

Tolbert, T. J. & Williamson, J. R. (1996). Preparation of specifically deuterated RNA for NMR studies using a combination of chemical and enzymatic synthesis. *J. Am. Chem. Soc.* **118**, 7929-7940.

Traub, P., Mizushima, S., Lowry, C. V. & Nomura, M. (1971). Reconstitution of ribosomes from subribosomal components. *Methods Enzymol.* **20**, 391-406.

Traub, P. & Nomura, M. (1968). Structure and function of *E. coli* ribosomes. V. Reconstitution of functionally active 30S ribosomal particles from RNA and protein. *Proc. Natl. Acad. Sci. USA* **59**, 777-784.

Treiber, D. K. & Williamson, J. R. (1995). A simple method for preparing pools of synthetic oligonucleotides with random point deletions. *Nucleic Acids Res.* **23**, 3603-3604.

Turchiniskii, M. F., Gus'kova, L. I., Khazai, I., Budovskii, E. I. & Kochetov, N. K. (1970). Chemical method of specific degradation of ribonucleic acids with selectively removed bases 3. Cleavage of phosphoester bonds in ribose 2- and 3-phosphates catalyzed by amines. *Mol. Biol. (USSR)* **4**, 343-348.

Turek, C. & Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA lignads to bacteriophage T4 DNA polymerase. *Science* **249**, 505-510.

Valegard, L., Murray, J. B., Stockley, P. G., Stonehouse, N. J. & Liljas, L. (1994). Crystal structure of an RNA bacteriophage coat protein-operator complex. *Nature* **371**, 623-626.

Van Ryk, D. I. & Dahlberg, A. E. (1995). Structural changes in the 530 loop of *Escherichia coli* 16S rRNA in mutants with impaired translational fidelity. *Nucleic Acids Res.* **23**, 3563-3570.

Van Stolk, B. J. & Noller, H. F. (1984). Chemical probing of conformation in large RNA molecules. *J. Mol. Biol.* **180**, 151-177.

Varani, G., Aboul-ela, F. & Allain, F. H.-T. (1996). NMR investigation of RNA structure. *Prog. Nucl. Mag. Reson. Spectroc.* **29**, 51-127.

Verwoerd, D. W. & Zillig, W. (1963). A specific partial hydrolysis procedure for soluble RNA. *Biochimica et Biophysica Acta* **68**, 484-486.

Vincze, A., Henderson, R. E. L., McDonald, J. J. & Leonard, N. J. (1973). Reaction of Diethyl pyrocarbonate with nucleic acid components. Adenosine. *J. Am. Chem. Soc.* **95**, 2677-2682.

Vladimirov, S. N., Babkina, G. T., Venijaminova, A. G., Ginautdinova, O. I., Zenkova, M. A. & Karpova, G. G. (1990). Structural arrangement of the decoding site of *Escherichia coli* ribosomes as revealed from the data on affinity labelling of ribosomes by analogs of mRNA derivatives of oligoribonucleotides. *Biochim. Biophys. Acta* **1040**, 245-256.

Wagner, G. (1993). Prospects for NMR of large proteins. *J. Biol. NMR* **3**, 375-385.

Weeks, K. M., Ampe, C., Schultz, S. C., Steitz, T. A. & Crothers, D. M. (1990). Fragments of the HIV-1 tat protein specifically bind TAR RNA. *Science* **249**, 1281-1285.

Weeks, K. M. & Cech, T. R. (1995). Protein faciliations of group I intron splicing by assembly of the catalytic core and the 5' splice site domain. *Cell* **82**, 221-230.

Weeks, K. M. & Cech, T. R. (1996). Assembly of a ribonucleoprotein catalyst by tertiary structure capture. *Science* **271**, 345-348.

Weeks, K. M. & Crothers, D. M. (1991). RNA recognition by tat-derived peptides: interaction in the major groove? *Cell* **66**, 577-588.

Weeks, K. M. & Crothers, D. M. (1992). RNA binding assays for tat-derived peptides: implications for specificity. *Biochemistry* **31**, 10281-10287.

Weiss, R. B., Dunn, D. M., Atkins, J. F. & Gesteland, R. F. (1987). Reading frame switch caused by base-pair formation between the 3' end of 16S rRNA and the mRNA during elongation of protein synthesis in *Escherichia coli*. *EMBO J.* **7**, 1503-1507.

Weitzmann, C. J., Cunningham, P. R., Nurse, K. & Ofengand, J. (1993). Chemical evidence for domain assembly of *Escherichia coli* 30S ribosome. *FASEB J.* **7**, 177-180.

Welch, J. B., Duckett, D. R. & Lilley, D. M. J. (1993). Structures of bulged three-way DNA junctions. *Nucleic Acids Res.* **21**, 4548-4555.

Welch, J. B., Walter, F. & Lilley, D. M. J. (1995). Two inequivalent folding isomers of the three-way DNA junction with unpaired bases: sequence-dependence of the folded conformation. *J. Mol. Biol.* **251**, 507-519.

Westhof, E., Romby, P., Romaniuk, P. J., Ebel, J.-P., Ehresmann, C. & Ehresmann, B. (1989). Computer modeling from solution data of spinach chloroplast and of *Xenopus laevis* somatic and oocyte 5 S rRNAs. *J. Mol. Biol.* **207**, 417-431.

Williams, D. P., Regier, D., Akiyoshi, D., Genbauffe, F. & Murphy, J. R. (1988). Design, synthesis and expression of a hyman interleukin-2 gene incorporating the codon usage bias found in highly expressed *Escherichia coli* genes. *Nucleic Acids Res.* **16**, 10453-10467.

Wimberly, B., Varani, G. & Tinoco, I., Jr. (1993). The conformation of loop E of eukaryotic 5S ribosomal RNA. *Biochemistry* **32**, 1078-1087.

Wintermeyer, W. & Zachau, H. G. (1970). A specific chemical chain scission of tRNA at 7-methylguanosine. *FEBS Lett.* **11**, 160-164.

Wintermeyer, W. & Zachau, H. G. (1975). Tertiary structure interactions of 7-methylguanosine in yeast $tRNA^{Phe}$ as studied by borohydride reduction. *FEBS Lett.* **58**, 306-309.

Woese, C. R., Winker, S. & Gutell, R. R. (1990). Architecture of ribosomal RNA: Constraints on the sequence of "tetra-loops". *Proc. Natl. Acad. Sci. USA* **87**, 8467-8471.

Woo, N. H., Roe, B. A. & Rich, A. (1980). Three-dimensional structure of *E. coli* initiator $tRNA^{fMet}$. *Nature* **286**, 346-351.

Wu, H., Jiang, L. & Zimmerman, R. A. (1994). The binding site for ribosomal protein S8 in 16S rRNA and spc mRNA from *Escherichia coli*: minimum structural requirements and the effects of single bulged bases on S8-RNA interaction. *Nucleic Acids Res.* **22**, 1687-1695.

Wurst, R. M., Vournakis, J. N. & Maxam, A. M. (1978). Structure mapping of 5'-^{32}P-labeled RNA with S_1 nuclease. *Biochemistry* **17**, 4493-4499.

Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York.

Wyatt, J. R., Chastain, M. & Puglisi, J. D. (1991). Synthesis and purification of large amounts of RNA oligonucleotides. *Biotechniques* **11**, 764-769.

Wyatt, J. R. & Walker, G. T. (1989). Deoxynucleotide-containing oligoribonucleotide duplexes: stability and susceptibility to RNase V_1 and RNase H. *Nucleic Acids Res.* **17**, 7833-7842.

Yamazaki, T., Lee, W., Revington, M., Mattiello, D. L., Dahlquist, F. W. & Kay, L. E. (1994). An HNCA pulse scheme for the backbone assignment of ^{15}N, ^{13}C, ^{2}H-labeled proteins: application to a 37-kDa *trp* repressor-DNA complex. *J. Am. Chem. Soc.* **116**, 6464-6465.

Yang, Y., Kochoyan, M., Burgstaller, P., Westhof, E. & Famulok, M. (1996). Structural basis of ligand discrimination by two related RNA aptamers resolved by NMR spectroscopy. *Science* **272**, 1343-1347.

Yonath, A. & Franceschi, F. (1993). Structural aspects of ribonucleoprotein interactions in ribosomes. *Curr. Opin. Struct. Biol.* **3**, 45-49.

Zakian, V. (1995). Telomeres: Beginning to understand the end. *Science* **270**, 1601-1607.

Zengel, J. M. & Lindahl, L. (1994). Diverse mechanisms for regulating ribosomal protein synthesis in Escherichia coli. *Prog. Nucleic Acids Res. Mol. Biol.* **47**, 331-370.

Zhang, P. & Moore, P. B. (1989). An NMR study of the helix V-loop E region of the 5 S RNA from *Escherichia coli*. *Biochemistry* **28**, 4607-4615.

Zimmerly, S., Guo, H., Eskes, R., Yang, J., Perlman, P. S. & Lambowitz, A. M. (1995). A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility. *Cell* **83**, 529-538.

Zimmerman, R. A. (1974). RNA-protein interactions in the ribosome. In *Ribosomes* (Nomura, M., Tissieres, A. & Lengyel, P., eds.), pp. 225-269. Cold Spring Harbor Lab. Press, Plainview, NY.

Zimmerman, R. A., Mackie, G. A., Muto, A., Garrett, R. A., Ungewickell, E., Ehresmann, C., Stiegler, P., Ebel, J.-P. & Fellner, P. (1975). Location and characteristics of ribosomal protein binding sites in the 16S RNA of *Escherichia coli*. *Nucleic Acids Res.* **2**, 279-302.

Zimmerman, R. A., Muto, A., Fellner, P., Ehresmann, C. & Branlant, C. (1971). Location of ribosomal protein sites on 16S ribosomal RNA. *Proc. Natl. Acad. Sci. USA* **69**, 1282-1286.

Zimmerman, R. A. & Singh-Bergmann, K. (1979). Binding sites for ribosomal proteins S8 and S15 in the 16S RNA of *Escherichia coli*. *Biochimica et Biophysica Acta* **563**, 422-431.

Zweib, C., Jemiolo, D. K., Jacob, W. F., Wagner, R. & Dahlberg, A. E. (1986). *Mol. Gen. Genet.* **203**, 256-264.

Zwieb, C. (1992). Recognition of a tetranucleotide loop of signal recognition particle RNA by protein SRP 19. *J. Biol. Chem.* **267**, 15650-15656.

Biography

Robert Todd Batey

b. Thousand Oaks, CA
May 20, 1968

Education:

August 1990 - present:	Ph.D. candidate, Biology Massachusetts Institute of Technology
September 1986 - June 1990:	B.S. in Chemistry University of California, Irvine Magna Cum Laude
	B.S. in Biological Sciences University of California, Irvine Magna Cum Laude

Research Experience:

April 1991 - present:	Graduate Research Assistant, Biology, M.I.T. Professor James R. Williamson (Dept. of Chemistry)
January 1989 - August 1990:	Undergraduate Research Assistant, Molecular Biology and Biochemistry, University of California, Irvine Professor Donald F. Senear

Teaching Experience:

September 1987 - August 1990:	General/Organic Chemistry Tutor, Tutorial Assistance Program, University of California, Irvine
June 1988 - August 1988:	Curriculum Coordination, Undergraduate Physics Laboratory, University of California, Irvine Professor William W. Heidbrink
September 1989 - June 1990:	Organic Chemistry Tutor, Department of Chemistry, University of California, Irvine,

Professor Mare Taagepera

September 1991 - December 1991:
Teacher's Assistant,
Biotechnology I Laboratory,
Massachusetts Institute of Technology,
Professors Philip Robbins and
Richard Young

January 1994 - June 1994: Teacher's Assistant,
General Undergraduate Biochemistry
Massachusetts Institute of Technology,
Professors Gene Brown and Peter Kim

Honors/Awards:

Phi Beta Kappa, May 1988

American Chemical Society, Orange County Section,
Outstanding Chemistry Major Award, June 1990

Special Service Award,
Department of Biological Sciences, U. C. Irvine,
June 1990

Keck Foundation Fellowship,
August 1990 - June 1991

Publications:

Senear, D. F., Batey, R. (1991). Comparison of Operator Specific and Nonspecific DNA Binding of Lambda *c*I Repressor: [KCl] and pH Effects. Biochemistry **30**, 6677-6688.

Batey, R. T., Inada, M., Kujawinski, E., Puglisi, J. D., Williamson, J. R. (1992). A general protocol for preparation of 13-C labeled RNA oligonucleotides for higher dimensional NMR study. Nucleic Acids Research **20**, 4515-4523.

Batey, R. T., Battiste, J. L., Williamson, J. R. (1995). Preparation of isotopically enriched RNAs for heteronuclear NMR. Methods in Enzymology **261**, 300-322.

Batey, R. T., Williamson, J. R. (1996). Interaction of the *Bacillus stearothermophilus* ribosomal protein S15 with 16S rRNA: I. Defining minimal RNA site. Journal of Molecular Biology **261**, 536-549.

Batey, R. T., Williamson, J. R. (1996). Interaction of the *Bacillus stearothermophilus* ribosomal protein S15 with 16S rRNA: II. Specificity determinants of RNA-protein recognition. Journal of Molecular Biology **261**, 550-567.

Batey, R. T., Cloutier, N., Mao, H., and Williamson, J. R. (1996). Improved large scale culture of *Methylophilus methylotrophus* for $^{13}C/^{15}N$ labeling and random fractional deuteration of ribonucleotides. Nucleic Acids Research **24**, 4836-4837.